SYSTAT®

Getting Started

Getting Started

Copyright

Copyright © 1989, SYSTAT, Inc.: all rights reserved worldwide. No part of this publication may be reproduced, transmitted, transcribed, stored in a retrieval system, or translated into any language or computer language, in any form or by any means, electronic, mechanical, magnetic, optical, chemical, manual or otherwise, without the express written permission of SYSTAT, Inc.

References

Excel is a trademark of the Microsoft Corporation.
SYSTAT and SYGRAPH are registered trademarks of SYSTAT, Inc.

SYSTAT, Inc.
1800 Sherman Avenue
Evanston, IL 60201-3793
708.864.5670 Tel.
708.492.3567 FAX

Preface

This is the first version of SYSTAT programmed completely on the Macintosh. Mark Bjerknes has completed a project of several years' duration on one of the largest Mac applications to date. Greg Mullins who joined SYSTAT after the project began, helped with numerous new features. Tom McHale worked with Greg to develop the Spin program.

This is also the first major revision of the SYSTAT and SYGRAPH documentation. MaryAnn Hill, who joined us this year as Director of Technical Services, is primarily responsible for this reorganization. I think it is the most usable manual we have produced.

The documentation could not have been completed without Erin Vang and Stacey Miceli. Erin wrote much of the material in *Getting Started* and produced page layouts. Stacey developed art work and contributed to the SYGRAPH manual.

Chris Soderquist, Pat Fleury, Mike Pechnyo, Amie Wilkinson, and others helped with quality control. Eve Goldman, Kris Yasutake, Stephanie Shishko, Lucy Saunders, and Tracy Koe supervised marketing and production.

Despite the technical difficulties of producing software of this magnitude for a machine like the Mac, we are all excited by the direction it has taken us. I want to thank all our Mac users and beta testers for suggestions which improved this version so enormously. I like our new look.

Leland Wilkinson
SYSTAT, Inc.
Evanston, IL

© 1989, SYSTAT, Inc.

Contents

1 Installing	2
2 Tutorial	16
3 Help	28
4 Data Editor	44
5 Data files	92
6 Redirecting output	110
7 Analysis	122
8 Graphing	140
9 Exploratory data analysis	164
10 Tips and shortcuts	194
Index	203

© 1989, SYSTAT, Inc.

Installing 1

Installing SYSTAT and SYGRAPH 4
New features with this version 4
Installing 5
Viruses 5
Technical notes 7
Time and space limits of procedures 7
SYSTAT files 7
SYSTAT supplements 8
Overview of documentation 9
The four volumes 9
Getting Started 9
DATA 9
SYGRAPH 10
SYSTAT 10
Notation 11
Macintosh basics 13

Tutorial 2

Open a file 18
Graphics 19
Scatterplot 22
Statistics 24
Descriptive statistics 24
Regression 29

© 1989, SYSTAT, Inc.

Help 3

Help 30
General help 30
Dialog help 33
Instant help 34
Menu items 34
Dialog boxes 35
Editor help 37
Command help 38
Information 41

© 1989, SYSTAT, Inc.

Data Editor 4

Introduction 46
Variables 46
Cases 46
Worksheet 46
Numeric and character variables 47

Entering data 48
Open the Data Editor 48
Maximum file size 50
Enter data 50
Variable names 51
Values 53
Save data 56

Editing data 59
Changing variable names 59
Changing data entries 59
Deleting cases and variables 60
Changing blocks of data 60
Selecting blocks 60
Copy, Cut, and Paste 62

Transforming data 65
Math... 65
Recode... 66
Operators and functions 68
Order of expression evaluation 76
Fill worksheet... 76
Filling variables with random data 77

Data Editor utilities 78
Find case... 78
Find next 78
Editor/Formats... 81
Decimals to show 82
Enter right/down 82
Matrix type 82
Help 82
Select... 82
Extract selected... 84
Remove select 84
Other ways to select data 85

Special data structures 86
Grouping variables 86
Aggregated data 88
Triangular matrices 90

© 1989, SYSTAT, Inc.

Data files 5

Saving data in files 94
Filenames 94
Saving data 94
Save... 94
Save as... 95
Single and double precision 96
Problems? 97

Printing data 98
Page Setup... 98
Print data... 98
Other ways to print data 98
Problems? 98

Importing data 99
Pasting data from the Clipboard 99
Reading text files 100
Reading Excel files 102
Problems? 103
Reading map files 103

Exporting data 104
Copying or Cutting into the Clipboard 104
Saving text files 104

Manipulating data files 106
Sort... 106
Rank... 108
Standardize... 109
Transpose... 109

© 1989, SYSTAT, Inc.

Redirecting output 6

Saving output in files 112
Graphs 112
Save graph as... 112
Viewing PICT files 112
Text 113
Notes... 113
Save selected text as... 114
Results to file as... 114
Data produced by analyses 115
Using saved data 115
Printing output 116
Page preview... 116
Graphs 116
Print Graph... 116
Results to printer 117
Text 117
Page formats... 117
Print selected text... 118
Results to printer 118
Data produced by analyses 119
Placing output in the Clipboard 120
Graphs 120
Text 120
Data produced by analyses 120

© 1989, SYSTAT, Inc.

Analysis 7

Setting up 124
Formats... 124
Show options in effect 126
Record preferences 127
Using dialog boxes 128
Selecting variables and functions 128
Functions 128
Keyboard shortcuts 128
Options 129
Cancel 129
Clear 129
Saving dialog box settings 130
Variable bundles 130
Setting up bundles 131
Switching bundles 132
Analyzing subgroups 134
By groups... 134
Select... 135
Other ways to select subgroups 136
Weighting cases 137
Weight... 137
Cross validation 138
Saving analyses after weighting 138
Non-integer weighting 138

© 1989, SYSTAT, Inc.

Graphing 8

Sizing and positioning graphics 142
The ruler tool 142
Positioning graphs by mouse 143
Resizing by mouse 144
Locating by typing 144
Resizing by typing 144
Example 145
Resize box 148
Moving graphs 148
Resizing graphs 148
Resizing example 148
Commands 149

Stacking and overlaying 150
Overlay, stack, and next 150
Overlay plot 150
Stack plot 150
Next plot 151
Movie... 151
Time series movie example 152

Options 157
Color... 157
Formats... 157
Global settings... 157

Drawing tools 158
Plot tools and colors... 158
Rectangle 159
Rounded-corner rectangle 159
Oval 159
Circle 160
Line 160
Text 160
Pointer 161
Colors 162
Hints 162

© 1989, SYSTAT, Inc.

Exploratory data analysis 9

Straightening scatterplots 166
SPLOMs 166
Straightening relationships 166
Outliers 170

Spinning 171
Spin 171
Home 172
X, Y, Z 173
Degrees 173
Labels 173
Background 173
Eye 174
Instant graphs 174
Brushing tools 174
Tutorial 174

Scatterplot brushing 182
Left arrow 183
Clicking on points 183
Clicking in white space 184
Marquee 184
Lasso 185
Subset analyses 185
Algebraic selection 186
Turning subset selections off 186
Extracting subsets 186

Tutorial 187

© 1989, SYSTAT, Inc.

Tips and shortcuts 10

All the windows are text editors 195
Redo last analysis 195
Save dialog settings 196
Bundles 196
Help 196
Scatterplot brushing tools 196
Command-key abbreviations 197
Command interface 197
Command logs 197
Command batches 197
Instant graphics and statistics 197
Variable names 198
Matrix cells 198
Example 198

© 1989, SYSTAT, Inc.

© 1989, SYSTAT, Inc.

© 1989, SYSTAT, Inc.

1 Installation

Installing SYSTAT and SYGRAPH 4
New features with this version 4
Installing 5
Viruses 5
Technical notes 7
Time and space limits of procedures 7
SYSTAT files 7
SYSTAT supplements 8
Overview of documentation 9
The four volumes 9
Getting Started 9
DATA 9
SYGRAPH 10
SYSTAT 10
Notation 11
Macintosh basics 13

© 1989, SYSTAT, Inc.

Installation 1

Overview

The first section of this chapter tells how to install SYSTAT and SYGRAPH on your Macintosh. Please take the time to look over our instructions even if you have installed SYSTAT products before.

The second section covers some technical details; you can probably just scan the section for now.

The third section, "Overview of documentation" gives an overview of this manual and the other three volumes, *DATA*, *SYGRAPH*, and *SYSTAT*. We discuss the organization, notation, and objectives of the manuals.

The final section, "Macintosh basics" is a brief survival guide to the Macintosh that might be useful if you have never used a Macintosh before or if our simple, step-by-step instructions don't seem so simple or step-by-step.

©1989, SYSTAT, Inc.

New features

If you have used previous versions of SYSTAT or SYGRAPH for the Macintosh, you will no doubt find this version very exciting.

Fully menu-driven operation

We now have a ***full*** implementation of the Macintosh point-and-click interface. You can do ***all*** of your work through menus and dialog boxes. SYSTAT and SYGRAPH have special menus with icons to remind you what all the different types of analyses and graphics look like. Both these menus can be ***torn off*** to sit ready anywhere on your desktop.

Dialog boxes remember your choices

All dialog boxes remember your choices from one procedure to the next, so if you want to—say—recompute a regression model without one variable, all you need to do is leave everything the way it was but remove one variable. If you want, you can save all of your dialog box settings from one session to the next.

Optional command interface

If you are an experienced SYSTAT or SYGRAPH user, you will be happy to know that you can still use commands if you want. We provide an optional command interface that even mouse-addicts will appreciate: you can use it to keep an audit trail of your work, to do repetitive analyses in batches, or to save time doing complex operations.

One integrated environment

All of SYSTAT's procedures ***and*** SYGRAPH are now one program. (No more switching modules.) SYSTAT and SYGRAPH share a Data Editor and data files. They also share utilities for manipulating and transforming data. Both programs are integrated to share one menu bar and to work simultaneously and ***interactively*** with the same data files and the same windows. Despite this integration, Macintosh SYSTAT is still fully compatible with SYSTAT on any other platform—minicomputer, MS-DOS versions, etc.

©1989, SYSTAT, Inc.

Links between analyses and graphics
In addition to all of SYSTAT and SYGRAPH being integrated, we have added HyperCard-like links between the Data Editor and the graph output window: *scatterplot brushing tools* let you see the relationships between your data and your graph visually, instantly. You can select and extract data visually, instantly. You can follow one case in the Editor through many plots in a SPLOM visually, instantly.

3-D plot spinning
You can spin three-dimensional scatterplots in real time. Animated 3-D plots could save you hours of computation and redrawing plots from different perspective. SYGRAPH even tells you what [Eye] setting to use to reproduce a perspective with a final, presentation quality graphic.

Instant help
You can get general help from a special Help window available through the Apple menu. You can get instant, context-sensitive help for any menu item or any dialog box or any button or any window—at a click of the mouse. You can get a quick review for any term or statistic in the output from an analysis.

Full color
If your machine supports color, you can draw graphs in full color.

Drawing tools
MacDraw-like drawing and writing tools let you customize your graphs in full color.

Installing

To install SYSTAT and SYGRAPH, insert Disk #1 into the floppy disk drive. Double-click the disk to see its contents. Then, double-click the Install application and follow the instructions it gives.

Viruses

Our products are extremely sensitive to viruses. One possible scenario: you install everything, you start SYSTAT up and everything goes fine, and you Quit. The next time—or a few times later—that you go to run SYSTAT, it won't start. You probably have a virus.

© 1989, SYSTAT, Inc.

Here are some ways to check whether you have a virus:

- Use virus-detecting programs such as Vaccine or Ferret; these detect many common viruses; contact your Macintosh dealer for help
- Check whether the file size of the SYSTAT application has increased since you installed (use **File/Get Info** from the Finder's menu)
- See whether the modification date in the **Get Info** screen has changed to a date when you ran SYSTAT

Our release disks do not contain viruses. All of our development and production machines are isolated physically and we take numerous precautions to ensure that they cannot become infected.

If you think your machine has a virus, use ResEdit or Interferon to investigate and eliminate viruses. Your Macintosh dealer can help you. Contact our Technical Support department if you have further questions.

© 1989, SYSTAT, Inc.

Time and space limits of procedures

Each procedure (including DATA) dynamically allocates memory for variables. The only way to discover the limits of the procedure is to try a large problem. In many procedures, the limits are on variables, not on cases. Almost every routine in SYSTAT can process as many cases as you can fit on your disk. The exceptions are some nonparametric and time series routines which store all the data in internal memory. The Data Editor can process the maximum number of variables allowed in a SYSTAT file. Not all the statistical routines can handle this many variables, however, because they use additional memory to process the analyses.

Processing time varies with machine. The input and output routines are slower than the statistical routines because SYSTAT must look for character variables and missing data and must convert numbers to characters before printing them. While this slows printing, it ensures that SYSTAT produces exactly the same results on any computer. In addition, the BASIC language provided by DATA is slow, since it interprets commands when they are executed. Again, the trade-off is that SYSTAT BASIC produces identical results on different machines. For most applications, DATA is slower than the statistical routines.

Disk space required varies with routines. Many routines use scratch files on the disk. Sorting, in particular, uses several extra files. The scratch files are stored in double precision, meaning at least 8 bytes for each number and 12 bytes for each character value. See Appendix I of the *DATA* manual for information on the file structure of SYSTAT files.

SYSTAT files

In addition to your data, SYSTAT files store variable names, the version of SYSTAT that created the file, and the file type. All versions of SYSTAT can read files created by previous versions of SYSTAT for the same operating system. The reverse is not always true. SYSTAT file types include rectangular (the usual), SSCP, covariance, correlation, dissimilarity and similarity. Not all SYSTAT procedures can process all types of files.

©1989, SYSTAT, Inc.

SYSTAT stores files in binary form so that it can process them much faster than it could raw data files. This means that you cannot directly transfer SYSTAT files between machines with different architectures and operating systems. You *can* transfer data files between machines through plain text (ASCII) files or with MACPC.

Note: Appendix II in the *DATA* manual, **SYSTAT file structure**, describes the internal format of SYSTAT files and provides FORTRAN subroutines to process them.

SYSTAT supplements

Supplementary modules available for SYSTAT and SYGRAPH provide highly specific data analysis or data management tools that are not part of the basic SYSTAT package. Current SYSTAT supplements include DESIGN (power analysis and expected mean squares), EzPATH (causal modeling), LOGIT (multinomial logistic regression), PROBIT, TOBIT, 2SLS (two-stage least squares), TESTAT (test item analysis, Rasch estimates), SETCOR (set correlation analysis), SIGNAL (signal detection analysis), and SURVIVAL(survival, reliability, and life tables analysis). Not all supplementary modules are available for every implementation of SYSTAT.

© 1989, SYSTAT, Inc.

Overview of documentation

The four volumes

These manuals are intended to serve both as reference guides—with concise, step-by-step instructions for using menus and dialog boxes—and as statistics and graphics handbooks. We hope that you will take the time to read each chapter carefully when you can, and we hope that you will be able to find information efficiently when all you have time to do is look up a procedure and go.

We have divided the documentation into four volumes: *Getting Started*, *DATA*, *SYGRAPH*, and *SYSTAT*.

Getting Started

This first book, *Getting Started*, is for just that: getting started with the basic program. You will learn how to use the SYSTAT Data Editor, how to handle files, how to save and print stuff, how to use the menus and dialog boxes, and how to do your statistical analyses and graphics.

Getting Started shows you how to use everything in the **File**, **Edit**, **Data**, **Goodies**, and **Editor** menus.

The next chapter is a **Tutorial** that will give you the basic feel of the program. We also present a data handling tutorial in the **Data Editor** chapter, a scatterplot brushing tutorial in the **Exploratory data analysis** chapter. And throughout this and the other manuals, we show you lots of examples that you can try—complete with data files, step-by-step instructions, and output.

This is the fun book. We show you how to spin plots. We show you how to use scatterplot brushing tools to link numbers to pictures. We show you how to get SYSTAT to do your work for you with instant graphs and statistics.

The last chapter, **Tips and shortcuts**, gives some hints for really flying through your work.

DATA

The *DATA* manual discusses the powerful DATA programming language that you can use for complex data manipulations. Most users will never need to open the *DATA* book, because you can accomplish so much with the SYSTAT Data Editor and the **Data** and **Editor** menus.

©1989, SYSTAT, Inc.

Your best bet is to set *DATA* on the shelf for a few weeks. After you are comfortable using SYSTAT and SYGRAPH and the Data Editor, flip through *DATA* and decide for yourself whether you need to tap DATA's resources.

SYGRAPH

The *SYGRAPH* manual shows you how to use SYGRAPH: in short, all the plots, graphs, and charts in the **Graph** menu. You should read *Getting Started* first, though—that's where we show you how to handle the menu and dialog box system, how to size, position, and overlay plots, how to show Movies of your plots, how to spin plots, and how to use scatterplot brushing tools.

The *SYGRAPH* manual has a separate chapter for each item of the **Graph** menu, and the chapters are organized according to graph type. Within each chapter, we show you first the major variations (the menu subitems) and then show how you can use the options to produce different types of graphs.

The *SYGRAPH* manual is more than just a software document telling what to click when. It's also a graphics handbook, containing much information about using graphs effectively. Take the time to *read* the manual.

SYSTAT

The *SYSTAT* manual shows you how to use SYSTAT: all the procedures in the **Stats** menu. Read *Getting Started* first to get the feel for the program and learn how to print, save, and handle data.

The *SYSTAT* manual has a separate chapter for each item of the **Stats** menu, and the chapters are organized alphabetically. Within each chapter, we have four sections: an introduction to the procedures, "Usage" instructions, technical details about "Computations," and "Examples" that you can try.

The *SYSTAT* manual, too, is more than a software document. It is a statistics handbook, with introductions for every class of procedures, advice, and examples. You will learn to use SYSTAT—and statistics—wisely by taking the time to study the *SYSTAT* book.

© 1989, SYSTAT, Inc.

Command references

The next three volumes, *DATA*, *SYGRAPH*, and *SYSTAT*, each have command references as their first appendix. These appendices give instructions and detailed syntax information for using the optional command interface.

Indices

Each volume has a comprehensive index of topics, menu items, dialog boxes, options, functions, etc.

Notation

Each chapter in every manual begins with an overview of the chapter, and a table of contents for the chapter appears on the facing page.

Headers

At the top of every page, we show the name of the section and the name of the chapter. Chapter titles always are shown in green in the outside corner.

Examples

In all the manuals we have numbered the examples within each chapter: 3.1, 3.2, 3.3, etc., where 3.2 is the second example in Chapter 3. For *SYSTAT* and *SYGRAPH*, we present an appendix of **Command equivalents for examples** that shows how to reproduce the examples using the optional command interface; the examples are organized by chapter and marked with their numbers in the chapters.

Instructions

When we show instructions for using menus and dialog boxes, we use bulleted lists of actions. Actions that are strictly required have solid bullets (•), and optional actions have open bullets (○).

- You must do this
- You might not need to do this

We show typed input (for the optional command interface) in a typewriter font:

```
THIS IS A COMMAND
```

All mouse instructions and command input appear in green, to make them easier to find.

©1989, SYSTAT, Inc

Menu items

The names of menus and menu items always appear in boldface, e.g. the **Corr** item from the **Stats** menu.

Some items in the **Stats** and **Graph** menus have subitems. For instance, **Bar** in the **Graph** menu has four subitems: **Bar**, **Percent**, **Range**, and **Stacked**. We sometimes abbreviate by saying "select **Bar/Range** from the **Graph** menu."

We also sometimes abbreviate regular menu items, e.g. **Data/Formats...** means "the **Formats...** item from the **Data** menu."

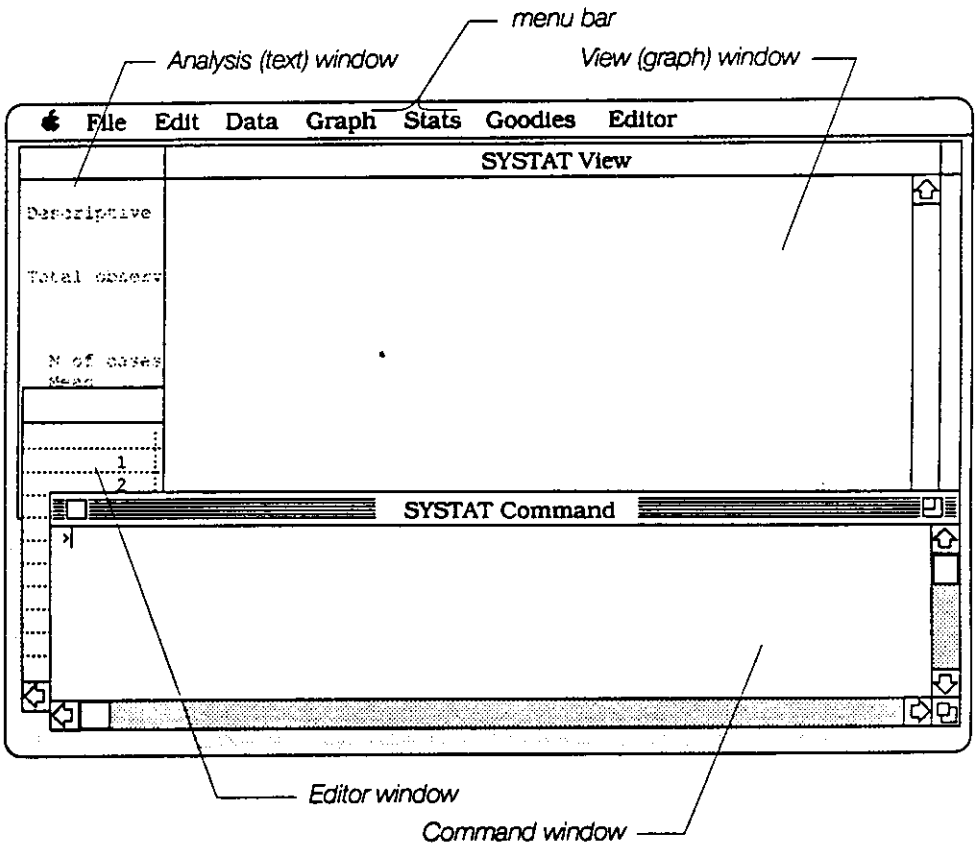
Options

We enclose the names of options—buttons, checkboxes, icons, etc.—in square brackets. For instance, you click the [Text] button in the **Save as...** dialog box if you want to save your data in a plain text file.

© 1989, SYSTAT, Inc.

If you have ever used a Macintosh before, you probably don't need to read this section. Most Macintosh software uses a consistent system of menus, windows, icons, and mouse techniques. The following does not tell all there is to know about the Macintosh interface, but it reviews the basic ideas of Macintosh usage. If any of the following paragraphs leave you confused, consult the *Macintosh System Software User's Guide* that came with your machine.

The top strip of the SYSTAT screen is a *menu bar* that contains *menus* of things you can do. Each menu contains *items* that accomplish certain tasks.



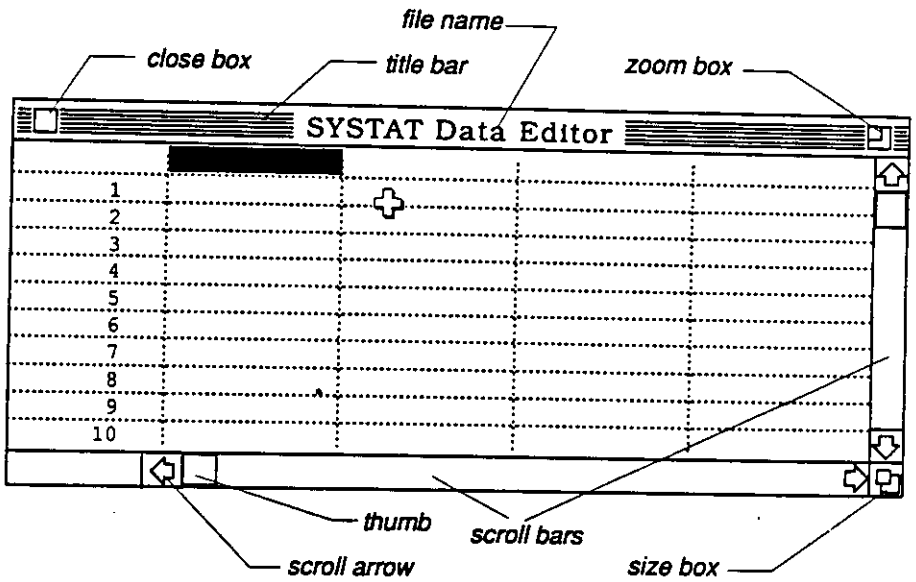
Macintosh software works with a system of *windows*. Many things can be going on at once, and you can switch around between very different types of tasks, because each task happens in its own window.

©1989, SYSTAT, Inc.

In SYSTAT, data are entered, manipulated, and viewed in the Data Editor window. Statistical analysis results are displayed in the Analysis window. Graphs are drawn in the View window. You can spin plots in the 3-D Window. If you choose to use the optional command interface, you type commands in the Command window.

At any given moment, one window is *active* and the others are dormant behind the current window. You can only work with the active window. To make a window active, click on it.

You can resize, relocate, and close windows. To resize them, click the *zoom box* in the upper right corner or *drag* the *size box* in the lower right corner. (To *drag* something, point the arrow at the item. Press and hold the mouse button, move the arrow to a new location, and release the mouse button.)



To move a window, drag its *title bar* at the top. To close a window, click the *go away box* or the *close box* in the upper left corner.

When a window contains more material than can fit in the window, the window *scrolls*—earlier information rolls off the top or off the left to make room for more. To scroll a window one line or column at a time, click the arrows at the extreme ends of the scroll bars. To scroll one page (windowful) at a time,click in the shaded *paging* area of the *scroll bars* at the left and bottom edges of the window. To make larger leaps, drag the *thumb* (the white box) to a new position.

© 1989, SYSTAT, Inc.

To use a menu item, pull down the menu and release the mouse button on the item you want. Some menu have Command-keystroke abbreviations listed to their right on the menu; you can access these items by holding the Command (⌘) or Apple key and typing the letter. For example, Quit is abbreviated ⌘-Q.

If you ever select an item from a menus, get a dialog box, and then decide you don't want to do the procedure, just click Cancel. You can then try something else.

You can use the mouse or cursor arrow keys to move around in the Data Editor and in the text windows. Press any arrow key once to move one row or column in that direction, and hold it down for repeated movements.

You can type numbers from both the numeric keypad and the top row of the standard keyboard.

©1989, SYSTAT, Inc

2 Tutorial

Open a file	18
Graphics	19
Scatterplot	22
Statistics	24
Descriptive statistics	24
Regression	26

© 1989, SYSTAT, Inc.

Tutorial 2

Overview

Welcome to SYSTAT! Although SYSTAT has more features than some popular mainframe statistical packages, you can begin using it right away. The chapters in this section introduce you to SYSTAT and the features, definitions and operations that are common to all modules in the SYSTAT system.

Before you begin, you should already know how to set up SYSTAT on your computer. If you have not yet done so, please read the installation and operating-system specific instructions in the previous chapter. There you will find how to make backup copies of the program, install SYSTAT on a hard disk drive, and launch SYSTAT.

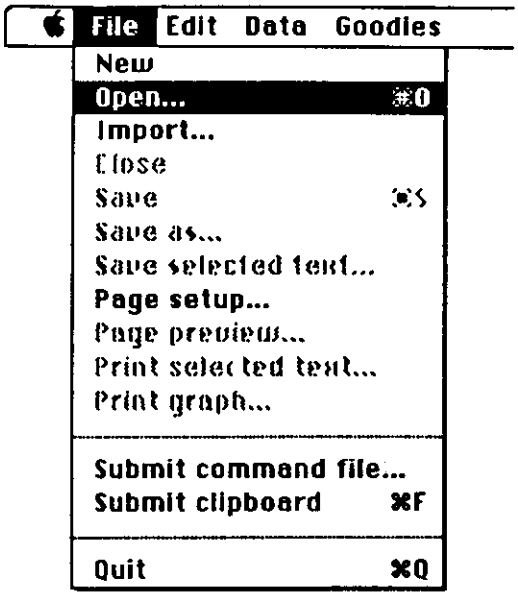
This tutorial takes you through a sample SYSTAT session. We use a file that has already been prepared. (The **Data Editor** chapter shows you step-by-step how to create your own SYSTAT data files.)

© 1989, SYSTAT, Inc.

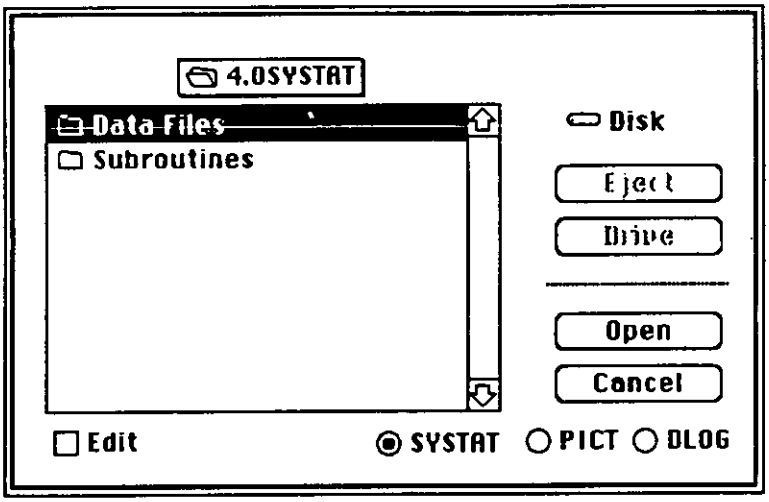
Open a file

Let's start with some statistical analyses. First, we open a file.

- Select **Open...** from the **File** menu

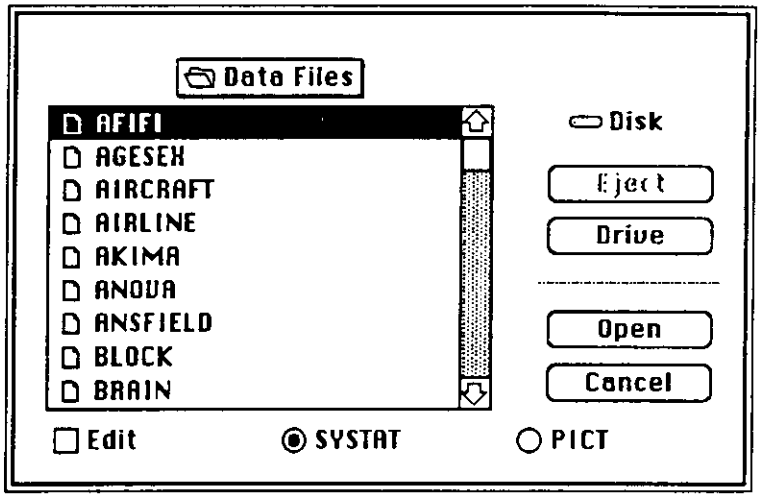


- Select the Data Files directory (double-click Data Files, or click Data Files and click OK)

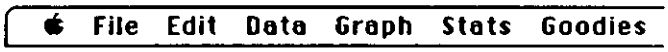


- Select the file USDATA
- Click Open

© 1989, SYSTAT, Inc.



Now, you should have a full menu bar.



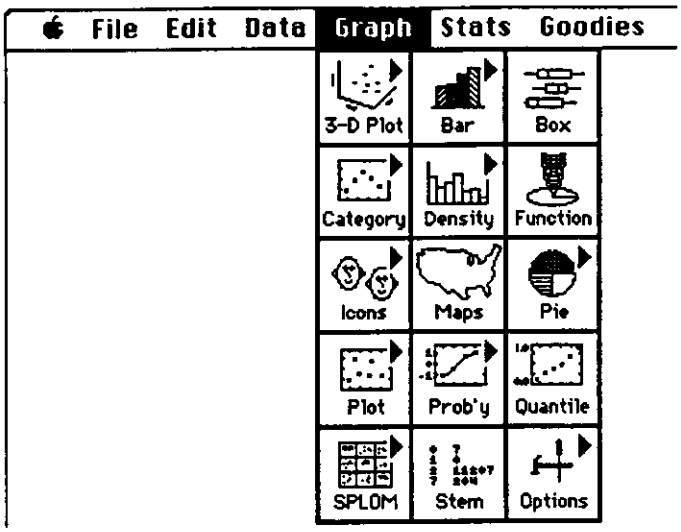
You are now ready to do statistical and graphical analyses. Since you should always look at your data before doing analyses, we are going to star with some graphics. First, we need to see the **Graph** menu.

The **Graph** and **Stats** menus are both *tear-off* menus. You can tear them off the menu bar and park them anywhere on your desktop. Your mouse saves on mileage, because you don't need to keep going up to the top of the screen and pulling through the menu to select an item—you just press, slide, and release.

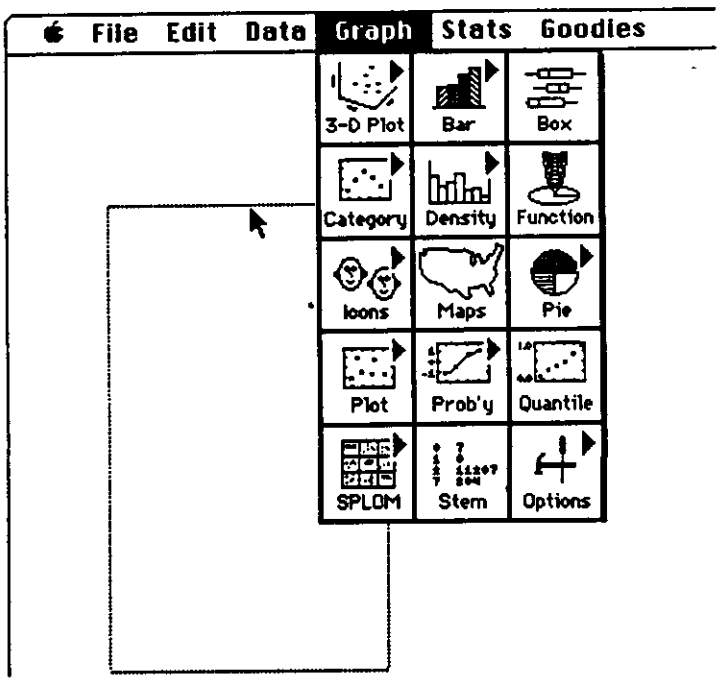
Graphics

For now, let's just tear off the **Graph** menu. Press and hold the mouse button on **Graph** in the menu bar. The **Graph** menu drops down. Keep holding the mouse button.

© 1989, SYSTAT, Inc.

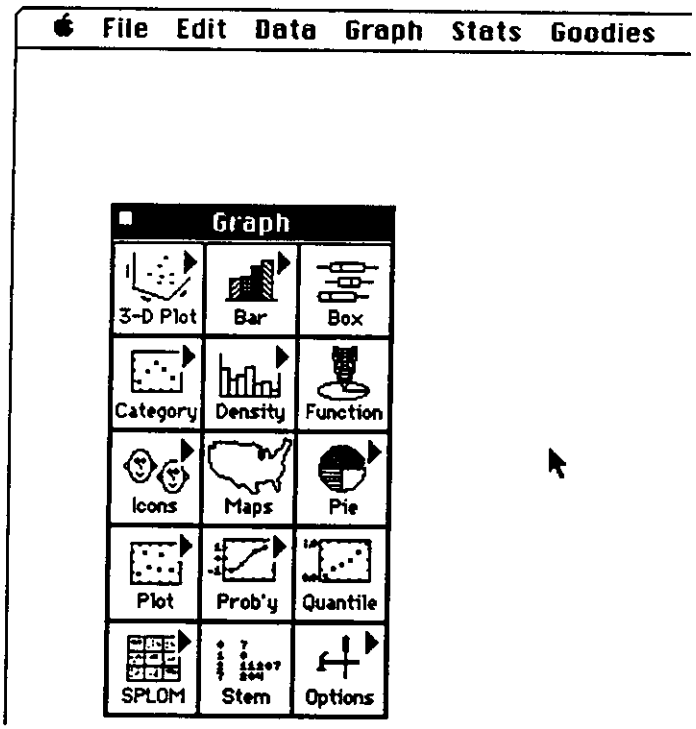


Now, slide the pointer through the menu and off one of its three sides. (Don't release the button yet!) Notice the marquee that trails your pointer.



Place this marquee where you want to park the menu and then—finally!—release the mouse button.

© 1989, SYSTAT, Inc.

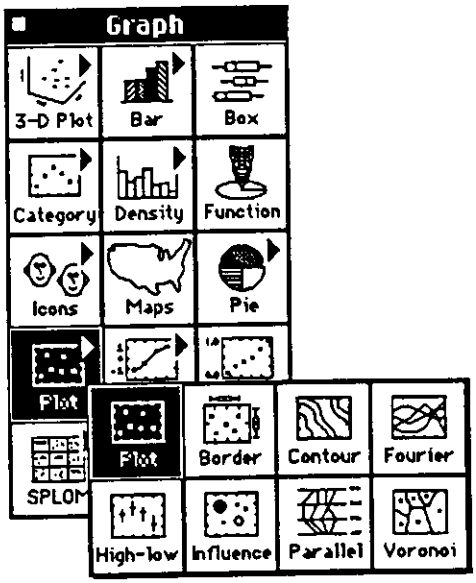


Voilá! Your very own torn-off **Graph** menu! Note that the menu looks a lot like a window. Sure enough, that white box in the upper left corner is a close box. The black strip is a title bar. You can close and move the menu just like you'd expect.

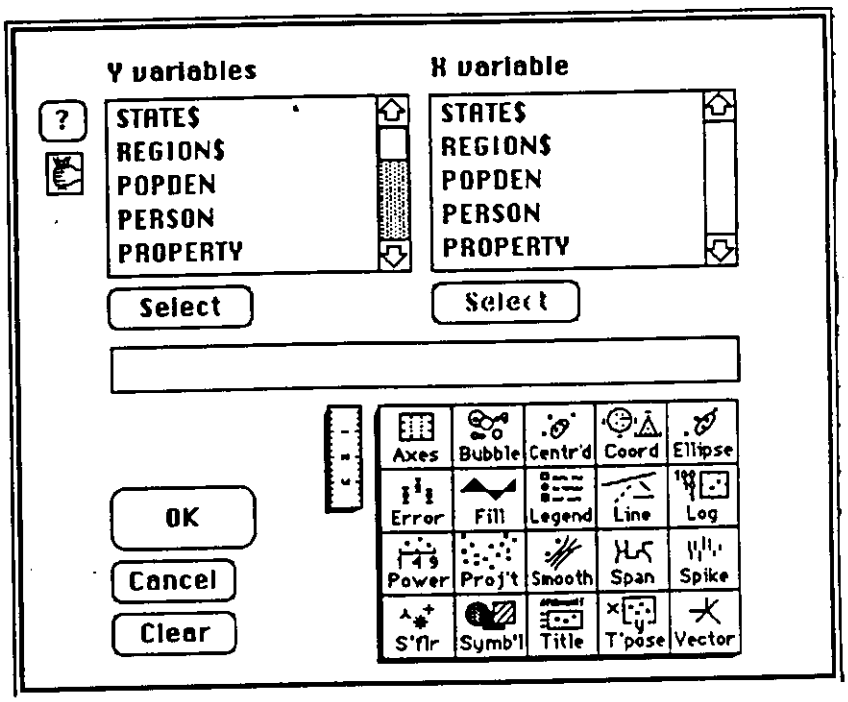
© 1989, SYSTAT, Inc.

Scatterplot

To use the menu, press and hold the mouse button on any of the items, e.g. **Plot**.



A submenu pops up offering variations of that type of plot, e.g. **Plot** , **Border, Contour, Fourier, High-low, Influence, Parallel**, and **Voronoi**. Slide the pointer to the item you want (**Plot**) and release the mouse button.

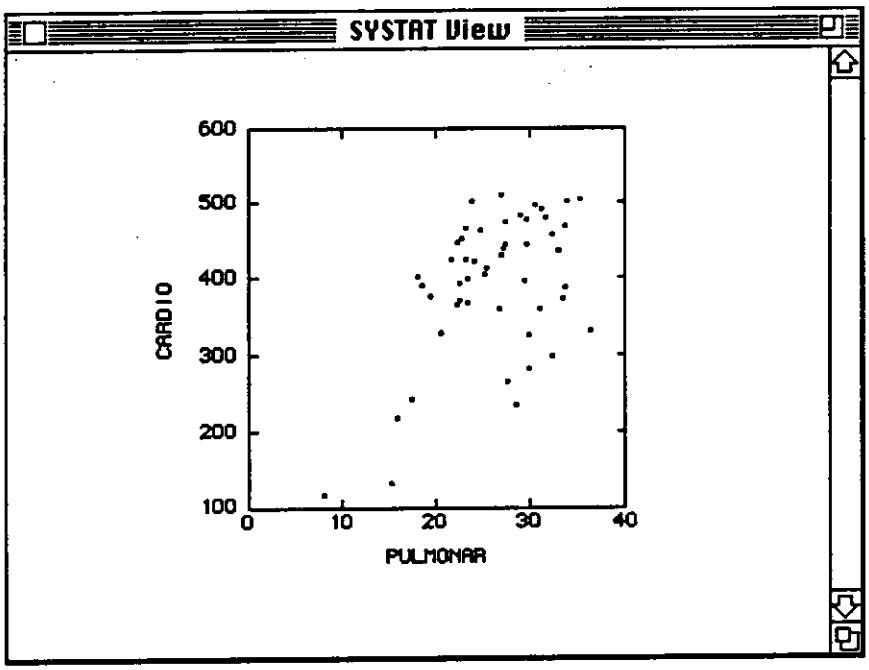


© 1989, SYSTAT, Inc.

Let's use this dialog box. We want a scatterplot of CARDIO and PULMONAR, two of the variables in the USDATA file.

- Select the Y (vertical axis) variable: scroll through the [Y variables] selection list until you find CARDIO; click on it and then click [Select]; or, double-click CARDIO
- Select the X variable: scroll through the [X variables] selection list to find PULMONAR, and either click it and click [Select] or double-click it
- Click OK

After a brief pause, a View window will appear on your desktop with the graph:



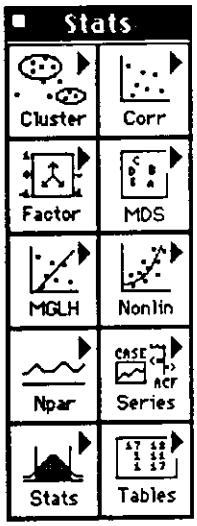
This plots the reported incidence of cardiologically related deaths versus pulmonary deaths for the 50 U.S. states in 1982. Each square represents a single state. Numbers within the plot indicate the number of states that have similar values for these two variables and plot at the same point.

The **Plot/Plot** dialog box you just used has many options in the lower right corner. Notice, however, that all you *have* to do is select Y and X variables.

© 1989, SYSTAT, Inc.

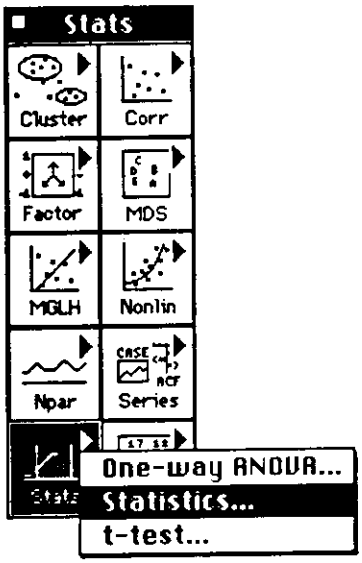
Statistics

Now let's try a SYSTAT procedure. First, tear off the **Stats** menu the same way you tore off the **Graph** menu. (You don't *have* to tear off the menus; we just think they're easier to use when they are sitting ready.)



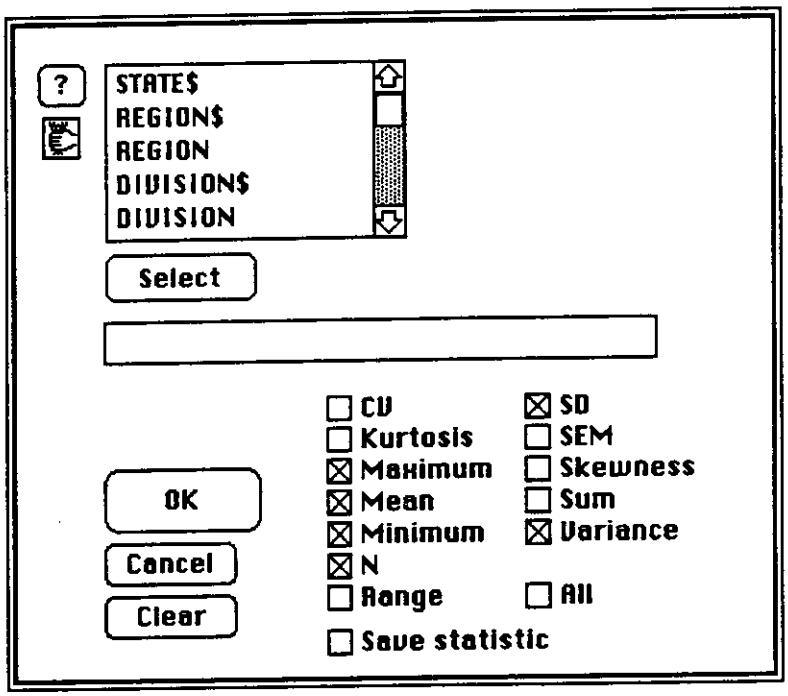
Descriptive statistics

Ready? Okay. Let's select **Statistics...** from the **Stats** item in the lower left corner of the Stats menu. (**Stats** is the only word that is both a menu and a menu item and—almost—a menu subitem; nothing else will be confusing this way. We promise!)



© 1989, SYSTAT, Inc.

The **Statistics...** dialog box looks like this:



Let's get some descriptive statistics on five of the USDATA variables. Notice that some of the statistic options are already checked: [Maximum], [Mean], [Minimum], [N], [SD], and [Variance]. These statistics are the *defaults*: the things you get automatically if you don't request something different.

- Select the variables: ACCIDENT, CARDIO, CANCER, PULMONAR, PNEU_FLU
- Click OK

© 1989, SYSTAT, Inc.

Total Observations: 50

	ACCIDENT	CARDIO	CANCER	PULMONAR	PNEU_FLU
N of Cases	50	50	50	50	50
Minimum	28.600	114.600	76.100	8.300	12.400
Maximum	85.800	508.700	244.000	36.300	32.400
Mean	44.312	398.530	178.410	26.486	21.038
Standard Dev	10.983	84.244	33.051	5.602	4.140

	DIABETES	LIVER
N of Cases	50	50
Minimum	3.400	6.200
Maximum	26.200	17.700
Mean	14.836	10.548
Standard Dev	3.668	2.745

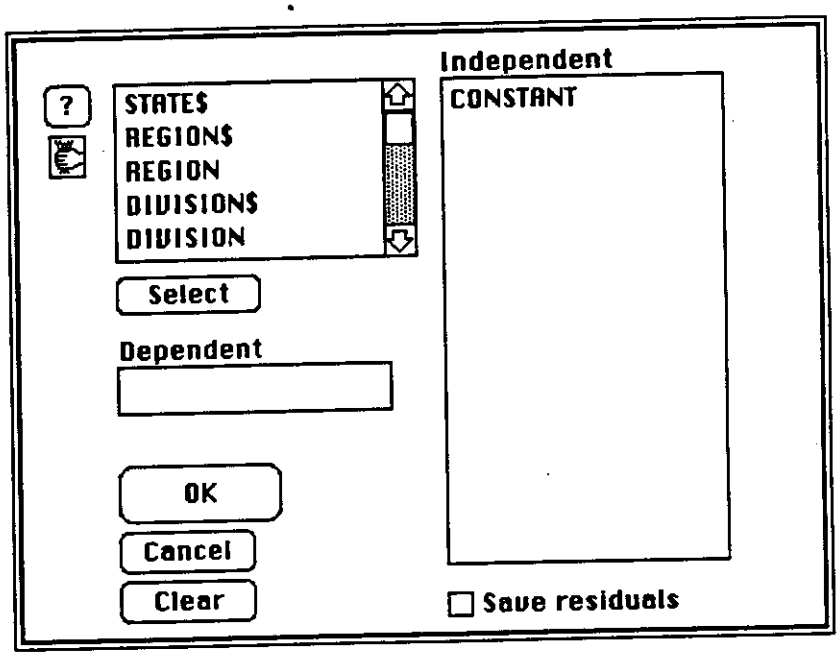
These are the default statistics. If you want other statistics or only a few of these, check the ones you want and uncheck any that you don't want.

Regression

Some graphic and statistical procedures have more subitems and options than others. You need not worry, however, about learning all the details before using a procedure. You can do almost any procedure simply, saving tricky options and variations for when you have more time and more experience.

Let's illustrate this point by using the most complicated menu item in the whole package: **MGLH**.

- Select **MGLH/Regression...** from the **Stats** menu



© 1989, SYSTAT, Inc.

Now let's do a simple regression of CARDIO on PULMONAR. You may notice that our data are not ideally suited for this model because the variances in CARDIO are not constant at all levels of PULMONAR and because there are dependencies across observations due to similar death rates in adjacent states. SYSTAT can be used to analyze more complex regression models suitable for data of this kind. Nevertheless, to keep this example simple, do the following:

- Select CARDIO from the [Dependent] variables selection list
- Click [Indep'dent] and select the independent variable and select PULMONAR
- Click OK

Notice that SYSTAT automatically included a CONSTANT term in the [Indep'dent] selection list. If you ever want to estimate a model without the CONSTANT (which we don't recommend), just select it with the mouse and press the Delete key.

Here is the output:

```
Dep Var:  CARDIO      N:   50    Multiple R:  .441    Squared Multiple R:  .194
Adjusted Squared Multiple  R:  .178     Standard Error of Estimate:       76.397

  Variable    Coefficient    Std Error     Std Coef Tolerance    T    P(2 Tail)

CONSTANT          222.932       52.716        0.000 1.0000000    4.229     0.000
PULMONAR            6.630        1.948        0.441 1.0000000    3.403     0.001

                                Analysis of Variance

   Source    Sum-of-Squares    DF  Mean-Square      F-Ratio       P

Regression       67598.890     1    67598.890       11.582       0.001
  Residual      280153.975    48     5836.541
```

© 1989, SYSTAT, Inc.

3 Help

Help 30
General help 30
Dialog help 33
Instant help 34
Menu items 34
Dialog boxes 35
Editor help 37
Command help 38
Information 41

© 1989, SYSTAT, Inc.

Help 3

Overview

Help for SYSTAT and SYGRAPH is never more than a click away. You can get on-screen help for any type of analysis or graphic, any menu item or subitem, any dialog box, any button, any check box, or any selection list. See the "Help" section.

If you choose to use the command interface, you can get help for the commands. See the "Command help" section.

Further, you can get information about any statistical term or title in the Analysis window output. See the "Information" section.

© 1989, SYSTAT, Inc.

You can get help for *anything* in SYSTAT and SYGRAPH.

General help: From the Apple menu, you can get general help for any type of analysis or graphing procedure along with any of the utilities for data selection, weighting, analysis by groups, and so on. See "General help," below.

Dialog help: You can get help for any dialog box by clicking the question mark button in the upper left of the box.

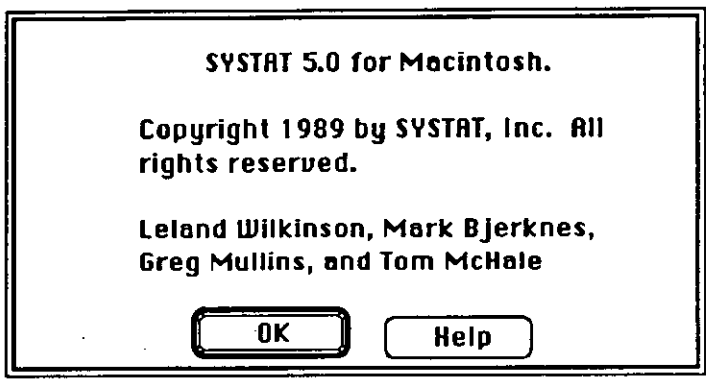
Instant help: Finally, you can get instant help for menu items and subitems and for options in dialog boxes by pressing the mouse button and then holding the question mark (?) key on your keyboard. Brief instructions pop up for the thing you are pointing at—a button or checkbox, a selection list, an icon, a menu, whatever! See "Instant help," below.

The **Editor** menu has an Help item that provides help for using the Data Editor. See "Editor help," below.

General help

For general information about SYSTAT and SYGRAPH procedures, use the general help feature in the Apple (🍎) menu at the far left of the menu bar. Select the first item in the menu, **About SYSTAT...**

The following appears. Click the [Help] button.


SYSTAT 5.0 for Macintosh.
Copyright 1989 by SYSTAT, Inc. All rights reserved.
Leland Wilkinson, Mark Bjerknes, Greg Mullins, and Tom McHale
OK
Help

© 1989, SYSTAT, Inc.

Help

The text window shows an introduction. For further topics, you can do one of three things:

1) Scroll through the vertical list of topics at the right. Click the topic you want, and you will see information for that topic.

2) Scroll through the horizontal list of icons at the bottom. Each icon represents an item in the **Stats** or **Graph** menu. Click the icon for the procedure or type of graph you want, and you will see information for that menu item.

3) Scroll through the text window itself. Topics are organized alphabetically. The two selection lists are shortcuts to the section pertinent to the item you select.

When you are done using help, you *must* click the close box on the Help window. You cannot resume working until you have closed Help.

Text list

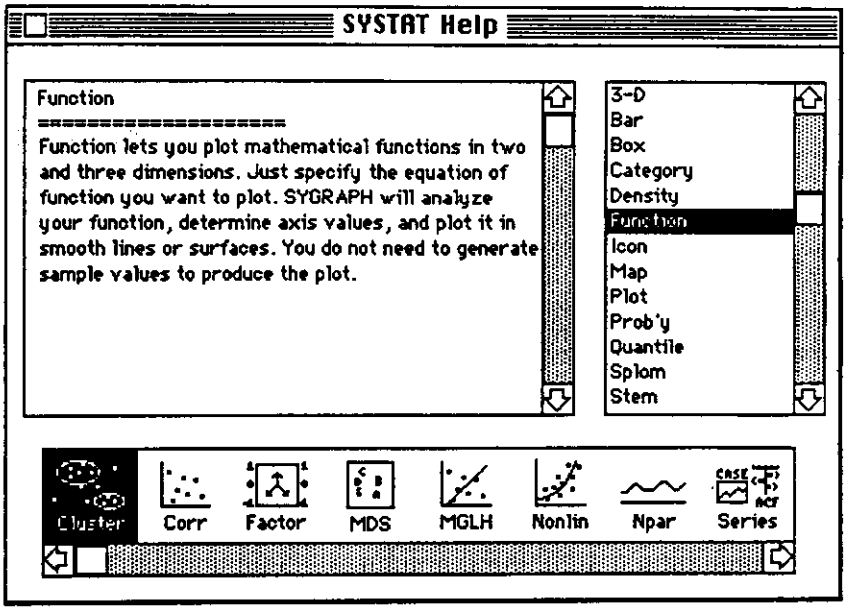
The text list shows a list of topics. If you click a topic, the text window jumps to its information about the topic.

Suppose you want information about function plots. Scroll down the list until you see the topic, and then click it. The text window jumps to its information about function plots.

© 1989, SYSTAT, Inc.

Help

You can also locate topics by typing the first few letters; the text list jumps to the first topic beginning with those letters.



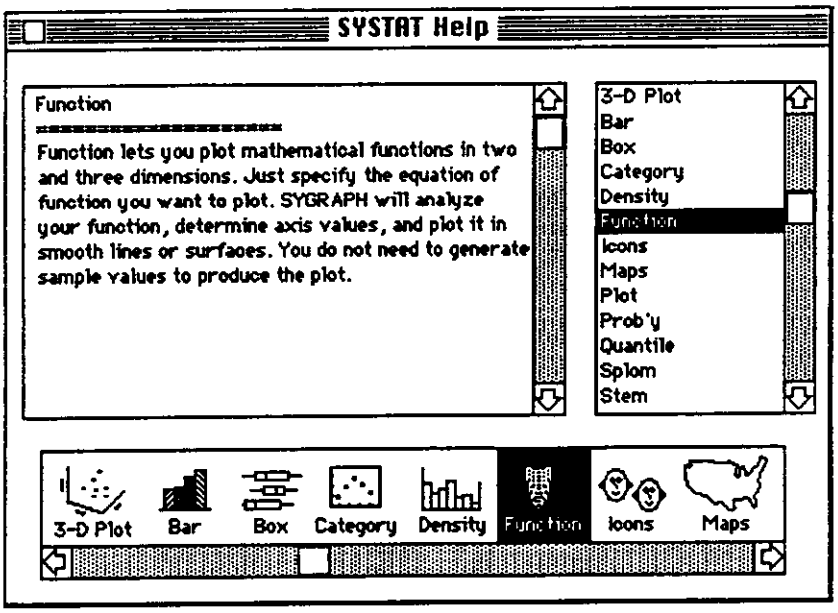
Icon list

The icon list shows the icons for **Stats** and **Graph** menu items, organized alphabetically, **Stats** first and then **Graph**.

You can scroll through the list to see **Graph** items. You could get the same information about function plots: just click **Function**, and the text window again jumps to its information about function plots. Notice also that the vertical list of topics jumps to the corresponding topic.

© 1989, SYSTAT, Inc.

Help



Text window

The topics are organized alphabetically in the text window. Feel free to browse—just use the scroll bars.

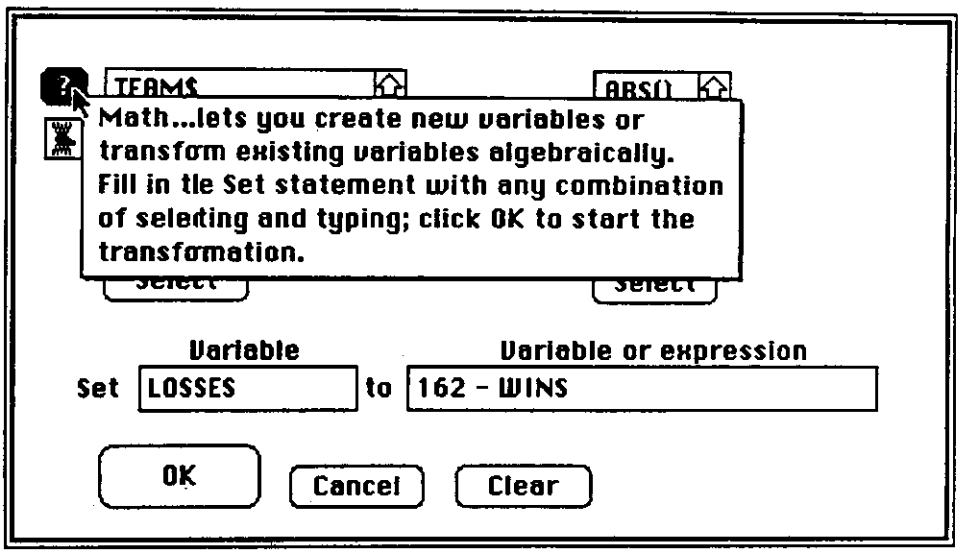
Dialog help

?

Every dialog box in the **Stats** and **Graph** menu and many **Data** and **Editor** menu dialog boxes (the ones that perform some procedure) have a question mark button in the upper left corner. Click the button to receive quick general instructions on how to use the dialog box.

© 1989, SYSTAT, Inc.

For instance, the help button in the **Editor/Math...** dialog box looks like the following:

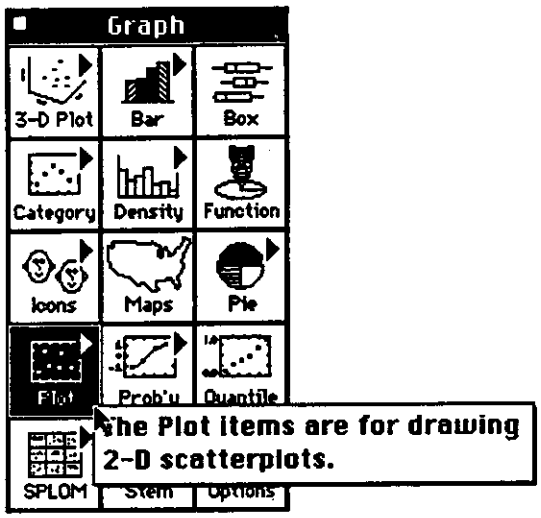


Instant help

By pressing with the question mark key, you can see brief help for any menu item or anything in a dialog box.

Menu items

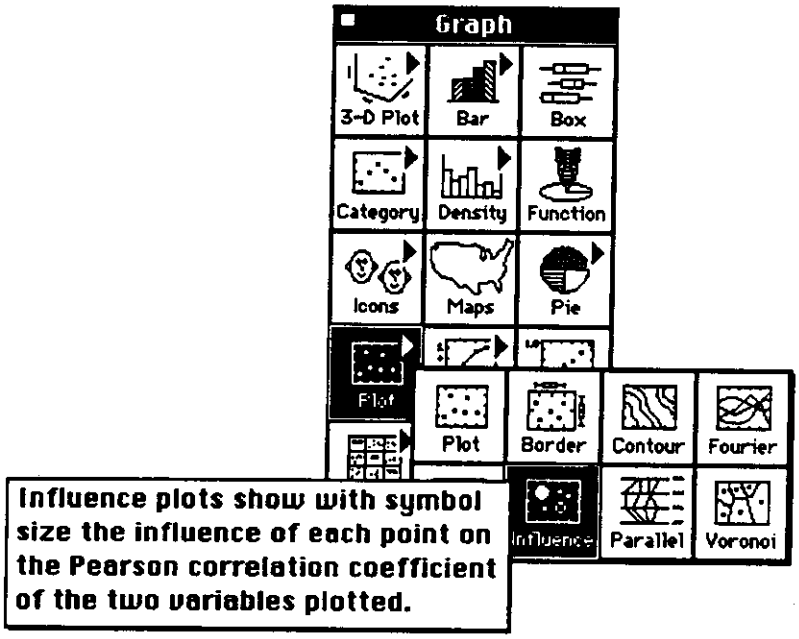
Suppose you're staring at the **Graph** menu and you just can't remember what **Plot** is for. Just point at the **Plot** item, press and hold the mouse button, and then hold the question mark (?) key on your keyboard. A quick reminder pops up:



You can also get help for any menu subitems. For instance, you can use ?-click to find out about **Plot/Influence**:

© 1989, SYSTAT, Inc.

Help



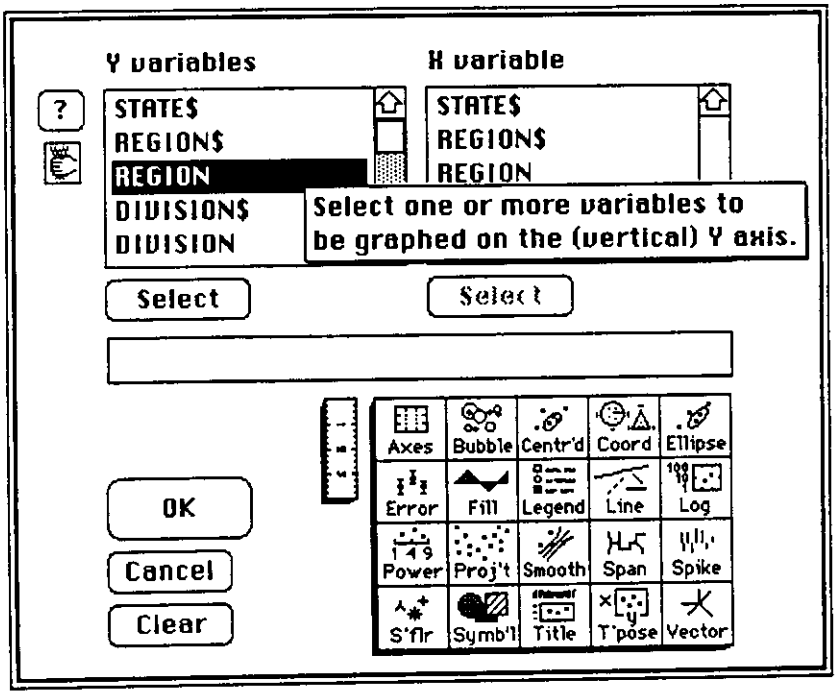
Dialog boxes

Use the ?-click technique to get information about items in dialog boxes. You can point at any item in the dialog box: a selection list, a button, a check box...

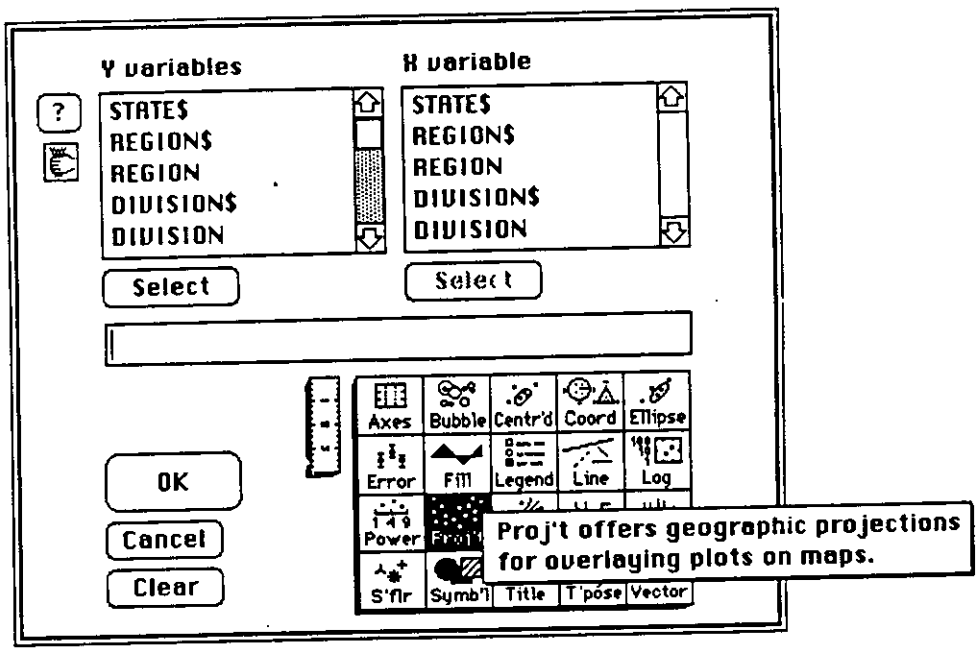
Suppose you decided to do a scatterplot but you weren't sure what to do with the [Y variables] selection list. Point inside the list, press the mouse button and hold it, and then press the question mark key.

© 1989, SYSTAT, Inc.

Help



Or, maybe you don't remember what [Proj't] is for. Point, press the mouse button, and press the question mark.



© 1989, SYSTAT, Inc.

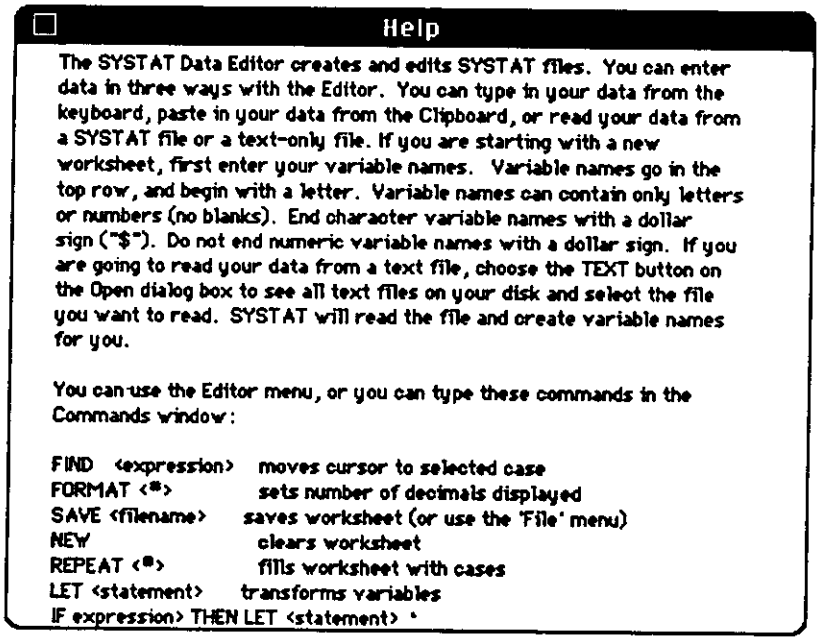
Help

Use ?-click whenever you need a quick reminder. If you need more complete help for something, try the help button (the question mark in the upper left corner of any dialog box) or the general help facility in the Apple menu.

Editor help

The **Editor** menu has an **Help** item. Use **Editor/Help** to get instructions for using the Data Editor.

- Select **Help** from the **Editor** menu


Help
The SYSTAT Data Editor creates and edits SYSTAT files. You can enter data in three ways with the Editor. You can type in your data from the keyboard, paste in your data from the Clipboard, or read your data from a SYSTAT file or a text-only file. If you are starting with a new worksheet, first enter your variable names. Variable names go in the top row, and begin with a letter. Variable names can contain only letters or numbers (no blanks). End character variable names with a dollar sign ("$"). Do not end numeric variable names with a dollar sign. If you are going to read your data from a text file, choose the TEXT button on the Open dialog box to see all text files on your disk and select the file you want to read. SYSTAT will read the file and create variable names for you.
You can use the Editor menu, or you can type these commands in the Commands window:
FIND <expression> moves cursor to selected case
FORMAT <#> sets number of decimals displayed
SAVE <filename> saves worksheet (or use the 'File' menu)
NEW clears worksheet
REPEAT <#> fills worksheet with cases
LET <statement> transforms variables
IF expression> THEN LET <statement>

For further help with the Editor, use the question mark buttons in the dialog boxes for specific **Editor** menu items or use HELP *command* in the Command window.

© 1989, SYSTAT, Inc.

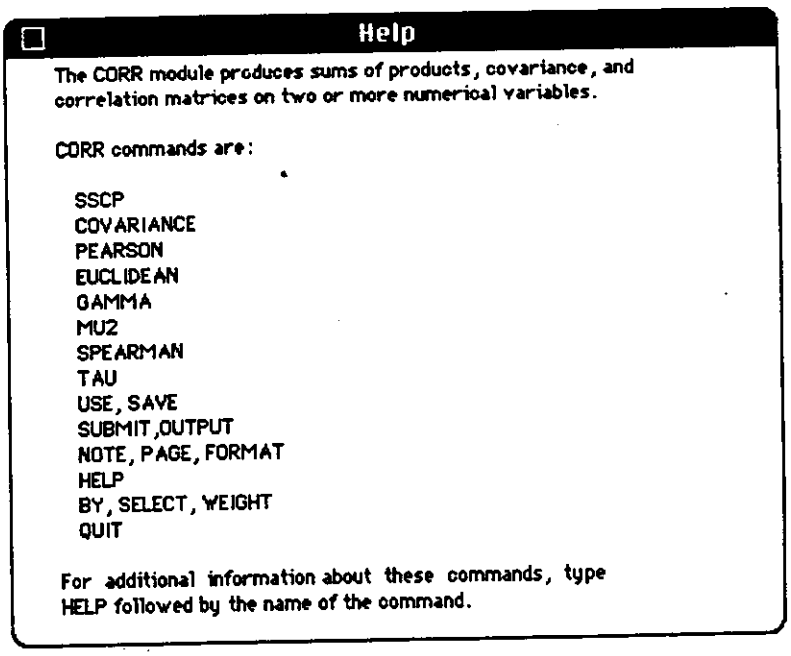
Command help

You can get help for the command interface by using the HELP command. First, you must open the Command window by selecting **Show command window** from the **Goodies** menu. For more information about the command interface, see the **Command reference** appendices in the *SYSTAT* and *SYGRAPH* volumes. Also see the appendix in the *DATA* volume for complete information about the DATA commands.

You can get both general and specific help by typing HELP alone or HELP with a command name , respectively. The command HELP without an argument produces a list of all the commands available in the current "module." If you need help for another module, switch to the module first by entering its name and then use the HELP command.

For example, if you are in CORR and use HELP, you will see the following:

```
HELP
```

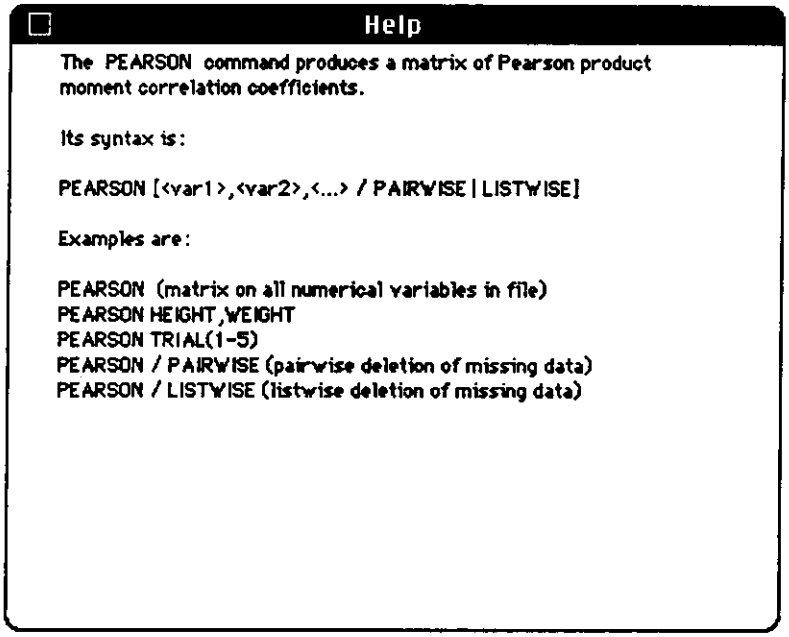


You can then get help for a specific command by typing the name of the command after HELP.

© 1989, SYSTAT, Inc.

You can get further help for the PEARSON command, for example, by typing HELP PEARSON.

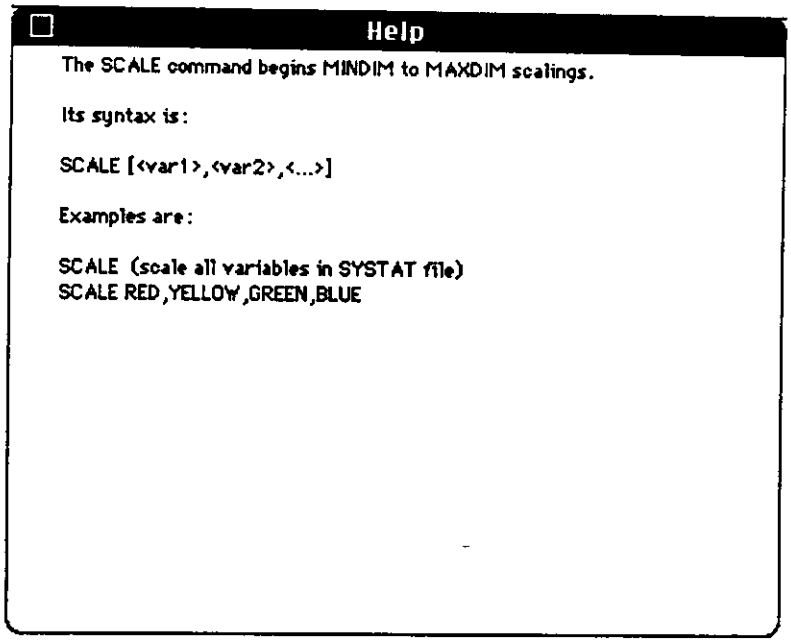
```
HELP PEARSON
```


Help
The PEARSON command produces a matrix of Pearson product moment correlation coefficients.
Its syntax is:
PEARSON [<var1>,<var2>,<...> / PAIRWISE | LISTWISE]
Examples are:
PEARSON (matrix on all numerical variables in file)
PEARSON HEIGHT,WEIGHT
PEARSON TRIAL(1-5)
PEARSON / PAIRWISE (pairwise deletion of missing data)
PEARSON / LISTWISE (listwise deletion of missing data)

© 1989, SYSTAT, Inc.

Suppose you are in CORR and need to get help for the MDS command SCALE. Just switch modules, and then use HELP.

```
MDS
HELP SCALE
```



You can get help for SYGRAPH by switching to the GRAPH "module" and then typing HELP or HELP *command*, where *command* is the command for which you need information.

You can type HELP *anywhere* in SYSTAT or SYGRAPH. Don't hesitate to use it. You can also take advantage of the NAMES and OPTIONS commands whenever you need to remember the variables, files, and options you are using.

© 1989, SYSTAT, Inc.

Information

Information from the **Goodies** menu gives information about any term or statistic you select.

- Select a term or statistic in the Analysis window output (use the mouse to highlight the text)
- Select **Information** from the **Goodies** menu (or type Command-I)

For example, **Stats/Statistics...** might produce some output like this:

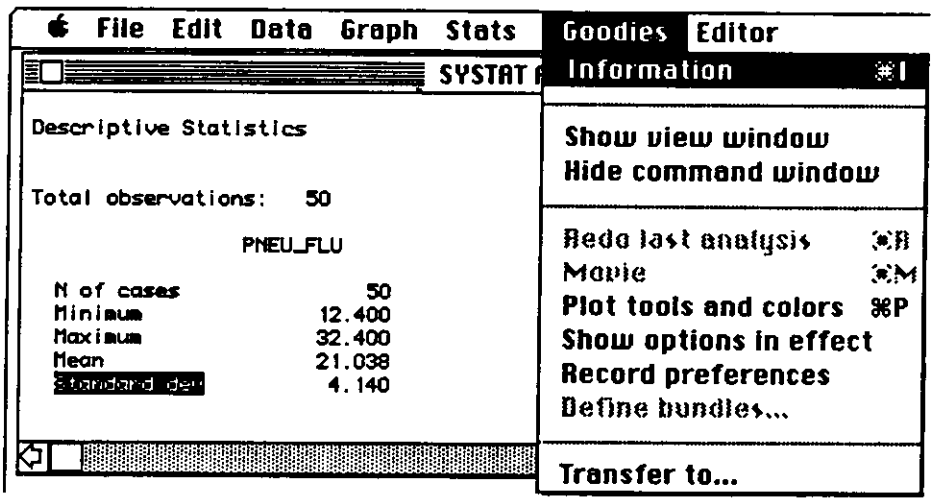
```
Descriptive Statistics

Total observations:   50

                PNEU_FLU

 N of cases             50
 Minimum            12.400
 Maximum            32.400
 Mean               21.038
 Standard dev        4.140
```

Can't remember what a standard deviation is? Select "Standard dev" in the analysis window, and use **Information.**



The Information window pops up to review the term:

© 1989, SYSTAT, Inc.

Information

Standard deviation, a measure of spread, is the square root of the mean squared deviation of values from the mean. Data sampled from a standard normal distribution have a standard deviation of one.

For series results, the standard deviation of the series is displayed. Standard deviation is the square root of the variance of the series data (the variable).

© 1989, SYSTAT, Inc.

© 1989, SYSTAT, Inc.

4 Data Editor

Introduction 46
Variables 46
Cases 46
Worksheet 46
Numeric and character variables 47

Entering data 48
Open the Data Editor 48
Maximum file size 50
Enter data 50
Variable names 51
Values 53
Save data 56

Editing data 59
Changing variable names 59
Changing data entries 59
Deleting cases and variables 60
Changing blocks of data 60
Selecting blocks 60
Copy, Cut, and Paste 62

Transforming data 65
Math... 65
Recode... 66
Operators and functions 68
Order of expression evaluation 76
Fill worksheet... 76
Filling variables with random data 77

Data Editor utilities 78
Find case... 78
Find next 78
Editor/Formats... 81
Decimals to show 82
Enter right/down 82
Matrix type 82
Help 82
Select... 82
Extract selected... 84
Remove select 84
Other ways to select data 85

Special data structures 86
Grouping variables 86
Aggregated data 88
Triangular matrices 90

© 1989, SYSTAT, Inc.

Data Editor 4

Overview

You can get data into SYSTAT four ways:

1) You can use the SYSTAT Data Editor
2) You can Copy or Cut data from another application
3) You can import text files from other applications
4) You can use SYSTAT's powerful DATA facilities

This chapter discusses the first two methods. It also shows how to edit data, transform variables, and search for cases.

See the next chapter, **Data files**, for instructions on importing.

See the *DATA* volume for information on the DATA command language for advanced data manipulations. You can also use DATA to input data from the keyboard (if you prefer) or to read fixed format ASCII files imported from mainframe applications.

This chapter combines a tutorial with reference information on the features of the Data Editor. We mark the tutorial with sidebars, so if you prefer to experiment first and read details later, you can follow the tutorial as it winds through the chapter.

© 1989, SYSTAT, Inc.

A data set is a collection of information. It contains various characteristics of a set of objects: the age, weight, and height of a bunch of people. The high, low, and closing prices of a group of stocks. The colors, flavors, and prices of a dozen soft drinks. The beginning, middle, and ending weights of patients at a diet clinic. The number of strikes, fouls, bases, and runs of baseball players.

For statistical analysis, we usually arrange this information in rows and columns. The rows comprise cases, or the objects we have measured. The columns comprise variables, or the characteristics on which we have measured the objects.

Variables

These characteristics—ages, prices, weights—are *variables*. A variable is a measure on which cases are assessed or an attribute describing them.

Cases

A *case* is one of the objects on which measurements are taken—one of the people, one of the stocks, one of the soft drinks, one of the baseball players.

Worksheet

A *worksheet* stores cases and variables in a rows-by-columns format. Say, for example, you are doing a medical study. You have a sample of 10 volunteers and are testing some new medications on them. You'll want to record some basic information about them: height, weight, age, sex, etc. You also need some way of distinguishing each volunteer—like name or chart number.

You might write all this information in a table like this one:

Name	*Height*	*Weight*	*Age*	*Sex*
Mildred	5'10"	140	33	F
Gertrude	6'2"	210	22	F
Clarence	5'2"	110	26	M
Melvin	5'11"	160	42	M

Here, the Name, Height, Weight, Age, and Sex columns are all variables. The Mildred, Gertrude, Clarence, and Melvin rows are all cases. Note that the *entire row* devoted to Mildred's characteristics (including her name) is one case. The *entire column* of names is one variable.

© 1989, SYSTAT, Inc.

You must set up data files like this for SYSTAT: with variables set up in vertical columns, and cases in horizontal rows. A rectangular arrangement of data like this one, with variables in columns and cases in rows, is called a *worksheet.*

Numeric and character variables

A variable can be *numeric* or *character*; that is, it can contain observations that are numbers (numeric) or that are some combination of letters, symbols, and numerals (character). In SYSTAT files, numeric variables contain decimal numbers only.

In our dataset, Weight and Age are numeric variables, because the values in those columns are numbers. Name and Sex are character variables, because their values are letters.

The tricky variable is Height: is it numeric or character? The values contain numerals, but they also contain the ' and " symbols for feet and inches, and they contain two fields: feet and inches. Numbers can only contain digits and an optional decimal point. Therefore, Height is a character variable.

You could change Height to make it a numeric variable: you could show Height by number of inches, or you could use the metric system (which uses only digits before and after a decimal point with a single unit, meters). Or, you could convert inches to tenths of feet.

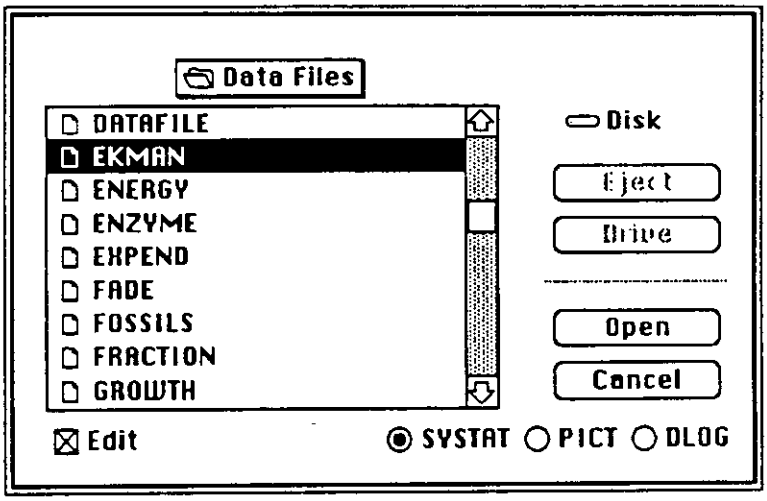
feet'inches"	*inches*	*meters*	*feet.tenths*
5'10"	70	1.778	5.833
6'2"	84	2.134	6.167
5'2"	62	1.575	5.167
5'11"	71	1.803	5.917

All but the first column could be numeric variables for SYSTAT.

© 1989, SYSTAT, Inc.

Open the Data Editor

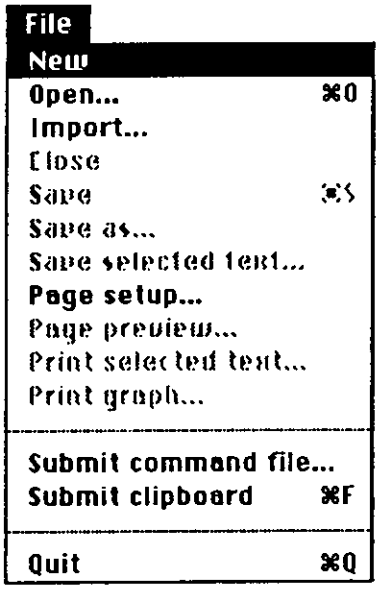
You enter data for SYSTAT by typing the worksheet into a Data Editor. You can open the Data Editor window two ways. You can open a preexisting file in the Data Editor by selecting **Open...** from the **File** menu and clicking [Edit] before clicking Open.



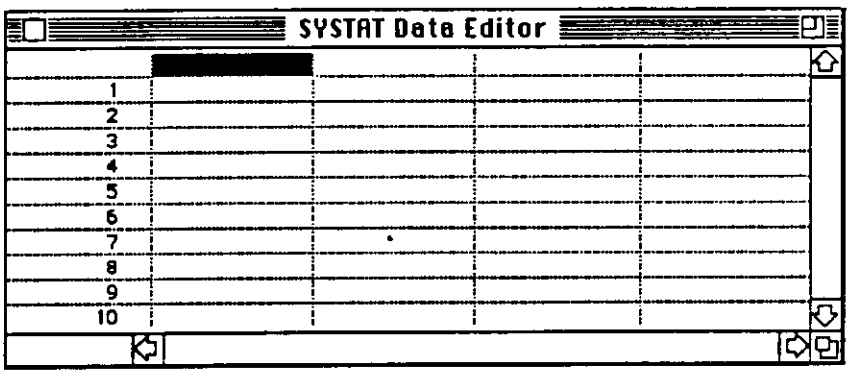
You can open an empty Data Editor window by selecting **New** from the File menu. Let's do that now. (You haven't started SYSTAT yet? Just double-click the SYSTAT application icon.)

© 1989, SYSTAT, Inc.

- Select **New** from the **File** menu



You should see an empty Data Editor window:



The Data Editor window is like any other window: you can move it, resize it, close it, and scroll around in it. You can **Cut**, **Copy**, and **Paste** to and from other windows and applications.

The Data Editor allows you to enter data into a worksheet, view data, edit data, and save data. Note, however, that *the Data Editor is not a spreadsheet.* It looks like a spreadsheet, but it differs from a spreadsheet in a number of important ways.

© 1989, SYSTAT, Inc.

The Data Editor is not a spreadsheet

First, you cannot store formulas in individual cells. You can use the **Editor** menu to do mathematical transformations, but the Editor only saves the results of the transformations—not the formula for making them.

For example: say you had entered two variables, A and B. Now you want a third column, A+B. Some spreadsheets would let you label the third column with the formula A+B, and the sums would fill the column. Then, if you changed the value in the third case for A, the third case for the A+B variable would automatically change to the new sum.

SYSTAT does *not* work this way. You would use **Math** from the **Editor** menu to "Set C to A + B," for instance, and C would be filled with the correct sums. However, if you then changed A and B values, C would not change. You would have to "Set C to A + B" again to replace the old sums with the updated sums.

Second, the Data Editor does not let you insert rows and columns like spreadsheets do. You can add variables and cases, but they are always placed after the last variable or column.

Finally, the Editor can handle as large a file as you can fit on your disk, while most spreadsheets are limited to a few thousand cells. It keeps the values in a file, not in memory; therefore, it is not limited by memory. Because of this, the Editor is somewhat slower than spreadsheets that work in memory.

Maximum file size

Your worksheet can contain up to 256 variables and as many cases as you can fit on your hard disk. If you had 20 megabytes of disk space available, you could store approximately 10,000 cases with 250 variables, which would make the worksheet 22 feet wide by 130 feet long. You could make a worksheet just as wide and a mile long if you had about 820 Mb of disk space. SYSTAT can handle it.

Enter data

To enter data into the worksheet, just select a cell and type in it. You can select a cell by clicking in it or by moving the cursor to it using cursor arrow keys. When you are done typing, press Return, Tab, or Enter. The cursor automatically jumps one cell to the right.

© 1989, SYSTAT, Inc.

(You may change things so that the cursor jumps down instead of to the right, if you prefer. Select **Formats...** from the **Editor** menu and click [Enter down]. If you want to make [Enter down] permanent, use **Record preferences** from the **Goodies** menu.)

Variable names

First you must enter variable names: labels for each of the columns you want to use. In other packages, variable names are sometimes called variable labels, column headings, or field headings. Variables names go in the top row of the worksheet.

Variable names can be up to eight characters (letters or numbers) long and must begin with a letter. You may only use letters and numbers—no symbols or punctuation marks. You may, however, use the underscore (_) character to represent a space, e.g. NEW_YORK.

You may type variable names in UPPER- or lower-case, but they are stored in UPPER-CASE.

We make these restrictions on variable names to ensure lateral compatibility: any SYSTAT data file or command file will work with SYSTAT on any machine—VAX, IBM-PC/compatible, UNIX, etc. (You do usually need to convert the file, but you won't ever have to rename variables and edit your command files.)

Character variables

Character variables (variables with values that are not numeric) *must* have a dollar sign ($) as their final character. The dollar sign does not count toward the eight character limit.

Note: you can change variable names in the Editor, but you may not change a character variable to numeric or vice versa. Thus, if you change the name of a character variable, the new name must also end with a dollar sign.

Subscripts

You may subscript numeric variables by placing the subscript numeral inside parentheses at the end of the variable name, e.g. X(1), X(2), ..., X(*n*). Subscripts are helpful because you could, for example, type STATS X(1–5) instead of selecting X(1), X(2), X(3), X(4), and X(5) individually.

© 1989, SYSTAT, Inc.

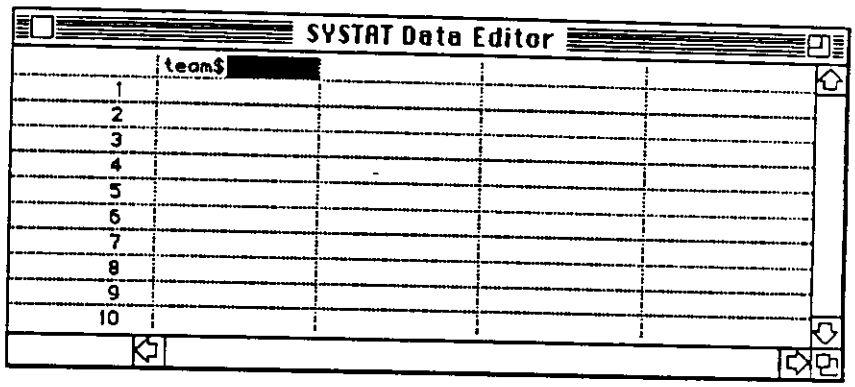
You may not subscript character variables. If a variable name has no $ at the end, SYSTAT assumes it to be numeric.

A set of subscripted variables is sometimes called an *array*. The number of variables in an array is its *dimension*. For example, the set of variables JUNK(1), JUNK(2), JUNK(3) is a three dimensional array.

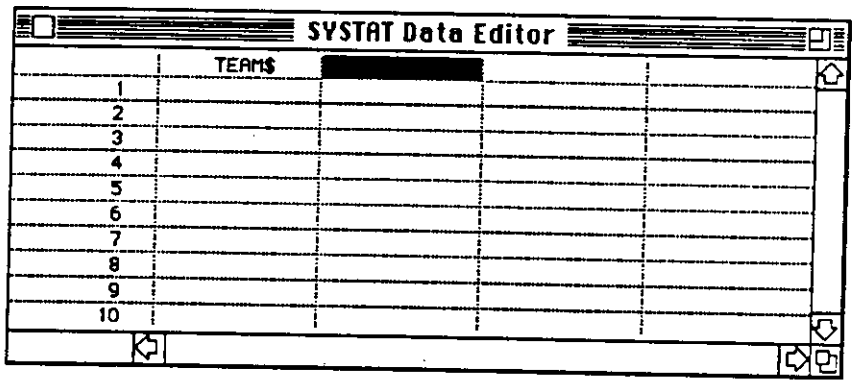
Entering names

To enter a variable name, select a cell and type the name and press Return, Enter, or Tab. The cursor jumps to the next cell to the right, and you can type the next variable name.

Try it out! Notice that the first cell of the worksheet is already selected. Type the first variable name, TEAM$, using either upper- or lower-case letters. Don't forget the dollar sign.

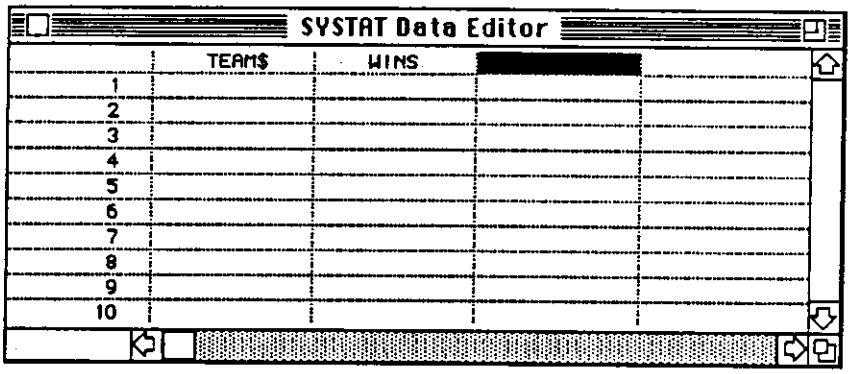


Now press return. The variable name is centered in upper-case, and the cursor jumps to the right:



© 1989, SYSTAT, Inc.

Now enter the second variable name, WINS. This is going to be numeric, so don't use a dollar sign. Don't forget to press Return (or Enter, or Tab).



Values

After you have entered variable names, you can enter values.

Numeric values

You can enter numeric values as integers or real numbers, with or without decimal places.

You can also use scientific notation. Exponents can have absolute magnitude less than 36. (That is, you can have values as large as $9*10^{35}$ and as small as $1*10^{-35}$.) Standard exponential shift notation applies:

$$2.9E5 = 2.9*10^{5} = 290{,}000$$
$$2.9E{-}5 = 2.9*10^{-5} = .000029$$

Thus, all of the following are valid numeric data values:

```
0
13579
1.0E15
.2E-5
123456789.1234
1.5e-12
2e25
```

© 1989, SYSTAT, Inc.

Character values

Character values may be up to 12 characters long. You may use upper- and lower-case; unlike variable names, though, character values are case-sensitive. That is, if you type lower-case characters, they remain lower-case.

If you need to use an apostrophe or single quotation marks as part of a character value, surround the whole value with double quotation marks. If you need double quotation marks as part of a value, surround the whole string with single quotation marks. You will not see the surrounding quotation marks in the stored value.

You can enter numeric values for a character variable *if* you surround the value with single or double quotation marks. SYSTAT treats such values as text; you cannot do numerical analyses with them.

All of the following are valid character values:

```
San Francisco
"Chuck's Place"
Three
'the "spot"'
```

Missing values

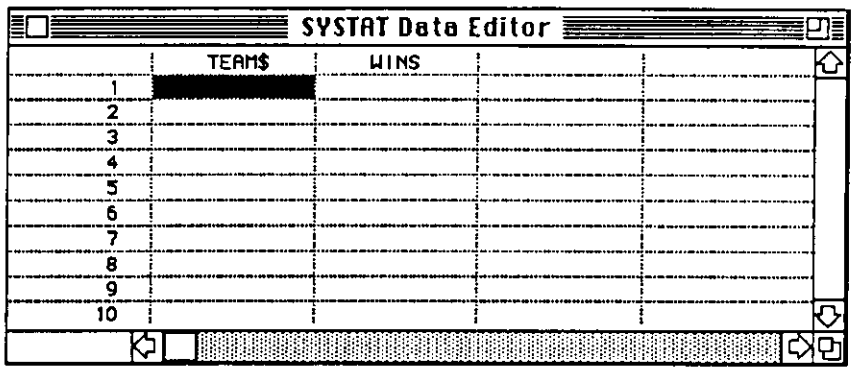
Suppose that you have a hole in your dataset. For example, Gertrude might refuse to divulge her weight. Or, perhaps a stock didn't get sold one day, or maybe somebody forgot to record its low price for the day. Perhaps one of the soft drinks is a test product that hasn't been named yet.

These would all be ***missing values***. You can't record a value in the table because you don't have one. You must, however, do something about that gap. On your scratch pad, you would probably just leave a spot blank, or maybe you'd make a hashmark to remind yourself that you didn't have a value (and that you didn't just skip that spot accidentally when you were writing things down).

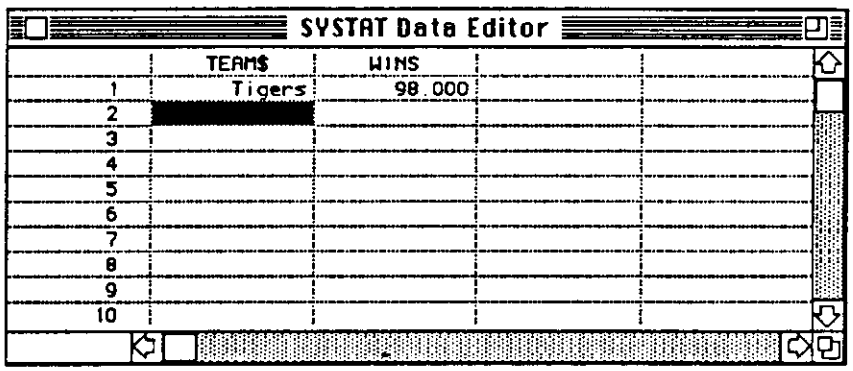
Enter a decimal point (period) by itself for a missing numeric value. Enter a blank space (press the Space bar and then Return) for a missing character value. If you don't enter one of these values and you leave a "hole" in your worksheet, the Editor will fill missing numeric values with a decimal point anyway.

© 1989, SYSTAT, Inc.

Let's enter some values. Select the cell immediately below TEAMS by clicking it or by moving to it with the cursor arrow keys.



Type the first team name, Tigers. The value is stored and the cursor jumps to the right. Now enter the first numeric value, 98. After you press return, the Editor should look like this:



Notice that the Editor is now ready for another TEAM$ name. Since you had filled the entire first case, SYSTAT assumed (correctly) that the next thing we would want to do is enter another team name.

© 1989, SYSTAT, Inc.

Enter some more data:

SYSTAT Data Editor

	TEAM$	WINS
1	Tigers	98.000
2	Orioles	67.000
3	Jays	96.000
4	Brewers	91.000
5	Yankes	89.000
6	Red Sox	78.000
7	Indians	100.000
8		
9		
10		

Notice that the Data Editor numbers your cases in the first column of the worksheet. In fact, CASE is a variable that is automatically present in any file.

Saving data

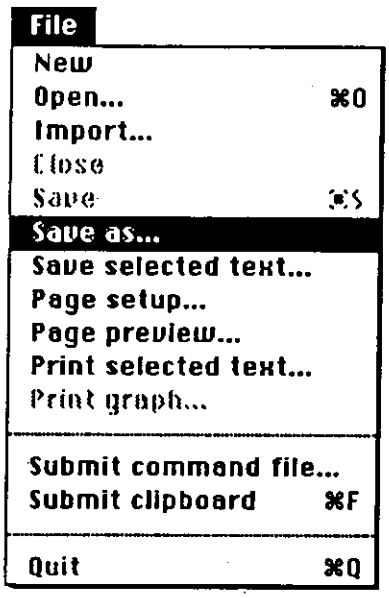
Two items in the **File** menu let you save data: **Save** and **Save as...** The first, **Save**, saves the data in the current filename or, if it is a completely new dataset, SYSTAT will ask you for a filename.

Say, for example, you had opened an existing file named TESTDATA and changed some values. If you then used **Save**, you would lose the old values which would be replaced by the new ones. Suppose you then discovered that your changes were all wrong and you wanted the old values back. It would be too late. If you had used **Save as...** instead and stored the new values in another file, such as TESTDATA2, all you would need to do is open the old file.

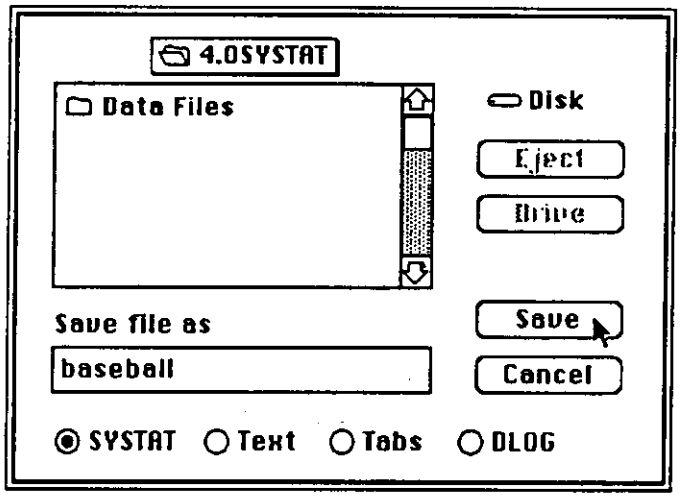
Let's save our data before we go any further. You never know when the power is going to fail or you're going to spill coffee on your keyboard. (Your keyboard may not survive, but at least you won't lose any data!)

© 1989, SYSTAT, Inc.

Just select **Save as...** from the **File** menu.



A dialog box asks for a filename. Type a name and click Save.



© 1989, SYSTAT, Inc.

SYSTAT reports that it has saved:

```
7 cases saved.
```

Now that you have given a filename, you can use **Save...** from the **File** menu to save any changes or additions you might make. **Save...** saves the data in the current file.

© 1989, SYSTAT, Inc.

As you have already seen, the Data Editor lets you change data values and variable names. Editing is the same as entering: click the cell, type the new value, and press Return. The old value disappears and the new one is stored.

Changing values in the Editor does not change the values in saved files. To change what is stored on your disk, use **Save** or **Save as...** from the **File** menu.

Changing variable names

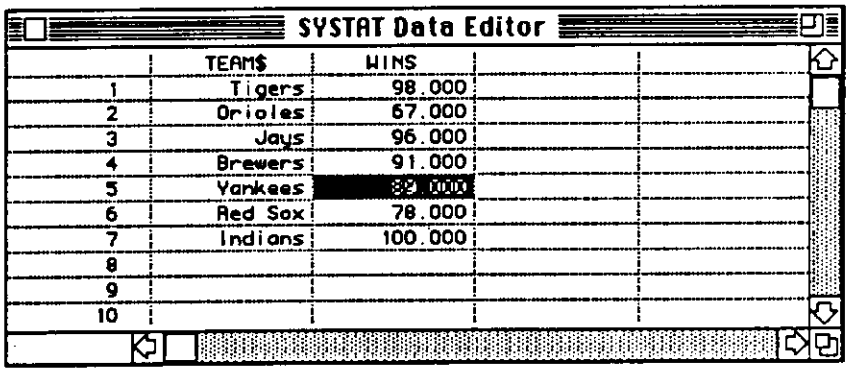
You can change a variables' name by clicking its cell and typing a new name. However, you cannot change a variable from numeric to character or vice versa. If you try, SYSTAT gives you an error message and leaves the name unchanged.

You should be careful to specify each variable's type correctly (by using a dollar sign $ for character variable names). If you make a mistake, you will have to **Delete** the variable and add a completely new variable to the right of the last variable in the worksheet. (We'll show you how to delete variables and cases soon.)

Changing data entries

Change data values the same way you change variable names: click the cell and type over the incorrect value. When you press Enter, the new value is stored and the old one disappears forever. (You cannot use Undo to change a cell back. You'll have to type it again.)

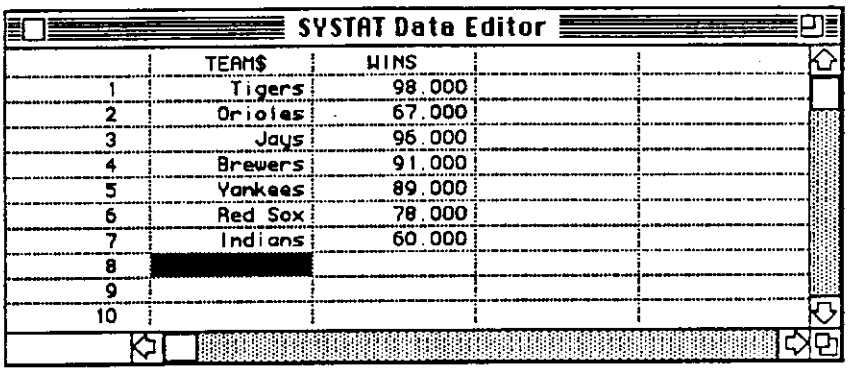
You probably noticed that we made a few mistakes in the baseball data. Click the "Yankes" cell and type the correct name, "Yankees." (You don't need to use the quotation marks.) Press return when you're done.



	TEAM$	WINS
1	Tigers	98.000
2	Orioles	67.000
3	Jays	96.000
4	Brewers	91.000
5	Yankees	82.000
6	Red Sox	78.000
7	Indians	100.000
8		
9		
10		

© 1989, SYSTAT, Inc.

The second mistake is more glaring. The Cleveland Indians? One hundred wins?! Click that cell and enter the correct total, 60.



Deleting cases and variables

To delete an entire row or column from the worksheet, use **Delete.**

- Select a row or column by double-clicking in its case number cell or variable name cell
- Select **Delete** from the **Editor** menu

You cannot undo **Delete.**

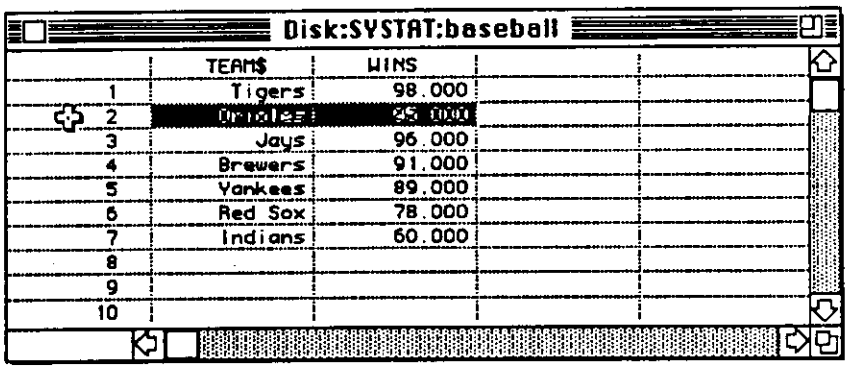
Changing blocks of data

You can select large blocks of data. Then, you can cut or copy the cells to the Clipboard and paste them into a new location of the worksheet, or in a different worksheet, or in a different window (like the Analysis window, that shows the results of statistical analyses), or in a different application.

Selecting blocks

Selecting rows

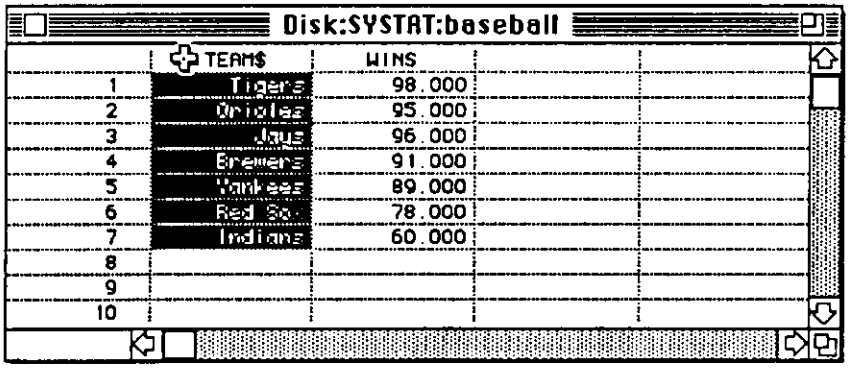
To select an entire row, just double-click in the cell containing that row's case number.



© 1989, SYSTAT, Inc.

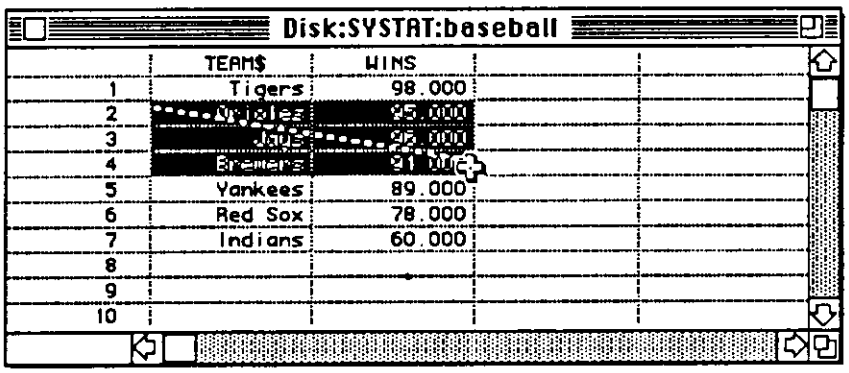
Selecting columns

To select an entire column, double click the cell containing the variable name.



Selecting blocks

To select a block of data, click the cell in one corner of the block you want, drag the pointer to the diagonally opposite corner of the block, and release the mouse button.



If the diagonally opposite corner is out of view, just drag outside the window, and the window will automatically scroll. Move the pointer back inside the window when you've scrolled far enough.

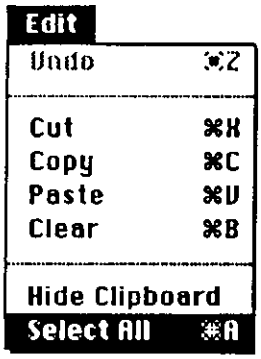
To select really large blocks of cells, click one corner cell, release the mouse button, and move to the diagonally opposite corner cell. (You may use the scroll bars as necessary.) Then, hold the Shift key and click the cell. You can add still more to your block by Shift-clicking in another cell.

© 1989, SYSTAT, Inc.

Note: you can only select rectangular blocks of cells. You cannot build irregular shapes by Shift-clicking; SYSTAT will always force your selection remain a rectangular block.

Selecting the entire window

You can select all of the data by selecting **Select all** from the **Edit** menu.



Copy, Cut, and Paste

Once you have selected a range of data, you can use **Copy** or **Cut** from the **Edit** menu to place it in the Clipboard.

Note: the following techniques also apply to text in the Command window and the Analysis output window.

Cut

You can cut a cell or range of cells from the worksheet. The cut data are removed from the worksheet and placed temporarily in the Clipboard (a storage place that carries from one file or one program to another, saving the *single most recent* cutting).

- Select a range of cells
- Select **Cut** from the **Edit** menu

Clear

Clear works the same as **Cut** except that the selected material is not sent to the Clipboard. Since nothing goes to the Clipboard, the contents of the Clipboard are unharmed, just as when you remove material using the Delete key.

Copy

Copying works like cutting, except that the selected data are not removed from the worksheet. A *copy* is placed in the Clipboard.

© 1989, SYSTAT, Inc.

- Select a range of cells
- Select **Copy** from the **Edit** menu

Cut and **Copy** both replace the previous contents of the Clipboard with new material.

You can display the Clipboard by selecting **Show Clipboard** from the **Edit** menu.

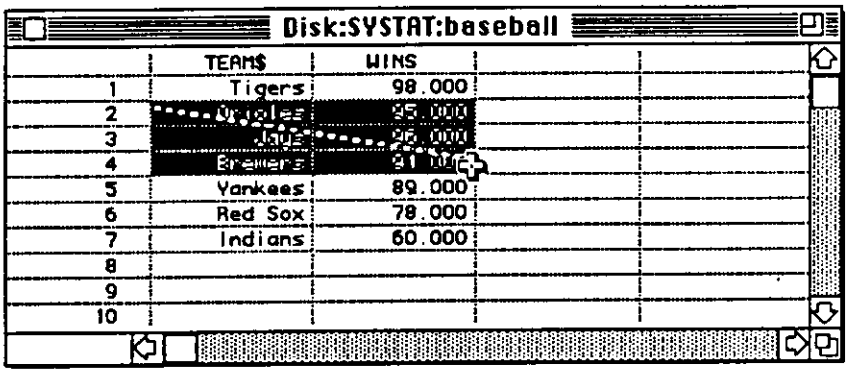
Paste

You may use **Paste** to place whatever is currently in the Clipboard somewhere else in the Data Editor (or in another window).

- Click the upper left corner of the area where you want the data
- Select **Paste** from the **Edit** menu

If you want, you can ignore the material in the Clipboard and consider it thrown away.

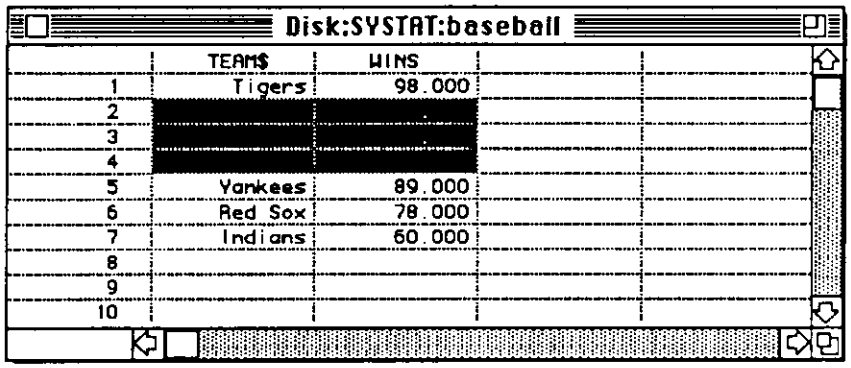
Let's try cutting some cells from the baseball data and pasting them back.



- Select the range as shown above
- Select **Cut** from the **Edit** menu

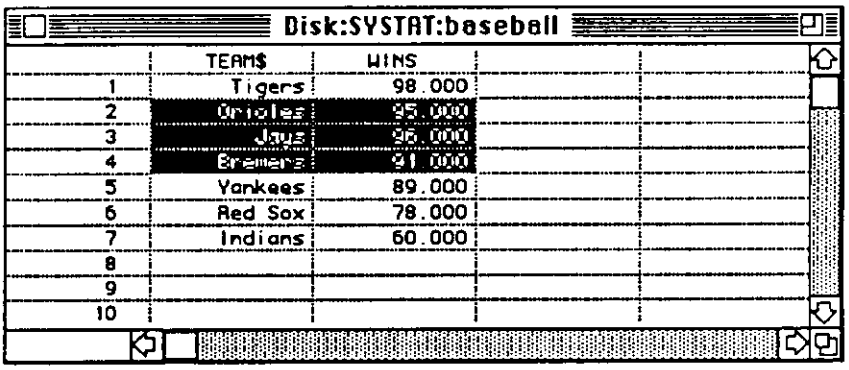
© 1989, SYSTAT, Inc.

SYSTAT fills the area with missing values:



Now, paste the data back in.

- Click the upper left cell in the empty area
- Select **Paste** from the **Edit** menu



We're back where we started.

© 1989, SYSTAT, Inc.

We said earlier that the SYSTAT Data Editor is not a spreadsheet. However, the **Editor** menu lets you do arithmetic with columns. **Math...** and **Recode...** let you do algebraic transformations. You can make numeric versions of character variables and vice versa. Finally, you can use **Fill worksheet...** to generate random values for variables.

Math...

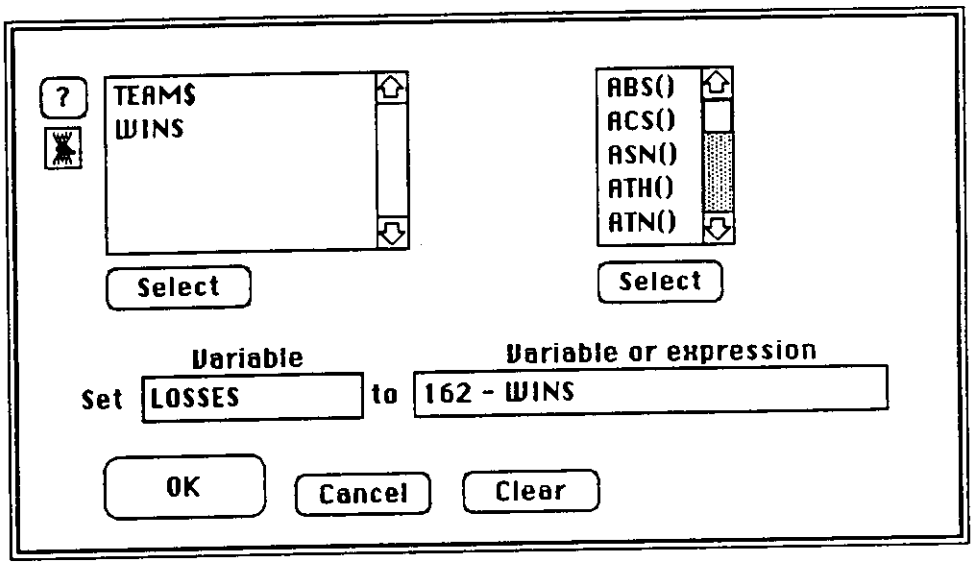
You can do algebraic transformations of data with **Math...** You can define (or redefine) one variable in terms of other variables and constants, using any of the operators, functions, and relations, which are discussed in the next sections. The dialog box presents a blank Set statement that you can fill in with any mathematically valid combination of variables, numbers, functions, and operators.

You can select variables and functions from the selection lists and type when necessary. Be sure to place the arguments of functions inside their parentheses. You can use the Tab key or the mouse to move from box to box.

In baseball there are 162 games in a season. In our file, the variable WINS contains the number of wins for each team in the Eastern Division of the American League for the 1986-87 baseball season. To enter the number of losses for each team we could name a new variable called LOSSES, calculate the number of losses by hand, and enter them cell by cell. Or, we could use **Math...**

- Select **Math...** from the **Editor** menu
- Fill in the Set statement: type LOSSES in the first box, click in the second box and type 162–WINS
- Click OK

© 1989, SYSTAT, Inc.



Watch the lower left corner of the Editor window. An odometer shows you the pace of SYSTAT's calculations. When the calculations are done, the new variable appears to the right of the last variable in the Editor.

Disk:SYSTAT:baseball

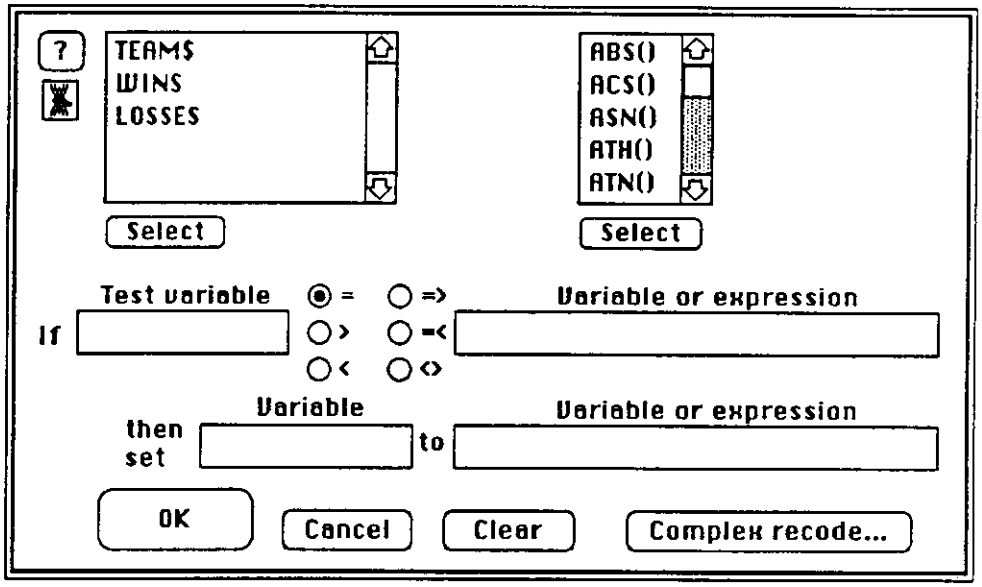
	TEAM$	WINS	LOSSES
1	Tigers	98.000	64.000
2	Orioles	67.000	95.000
3	Jays	96.000	66.000
4	Brewers	91.000	71.000
5	Yankees	89.000	73.000
6	Red Sox	78.000	84.000
7	Indians	61.000	101.000
8			
9			
10			

Recode...

Recode... lets you define variables in terms of others conditionally with if-then statements. The dialog box presents a blank if-then statement which you can fill in with any mathematically valid expressions.

Let's make a character variable showing which teams have extreme win-loss records. We want to make a LEVEL$ variable that says 'Extreme' for any team that has won or lost more than 95 games. For this, we need to use the logical operator [Or]. Note, however, that [Or] is not available in the **Recode...** dialog box:

© 1989, SYSTAT, Inc.

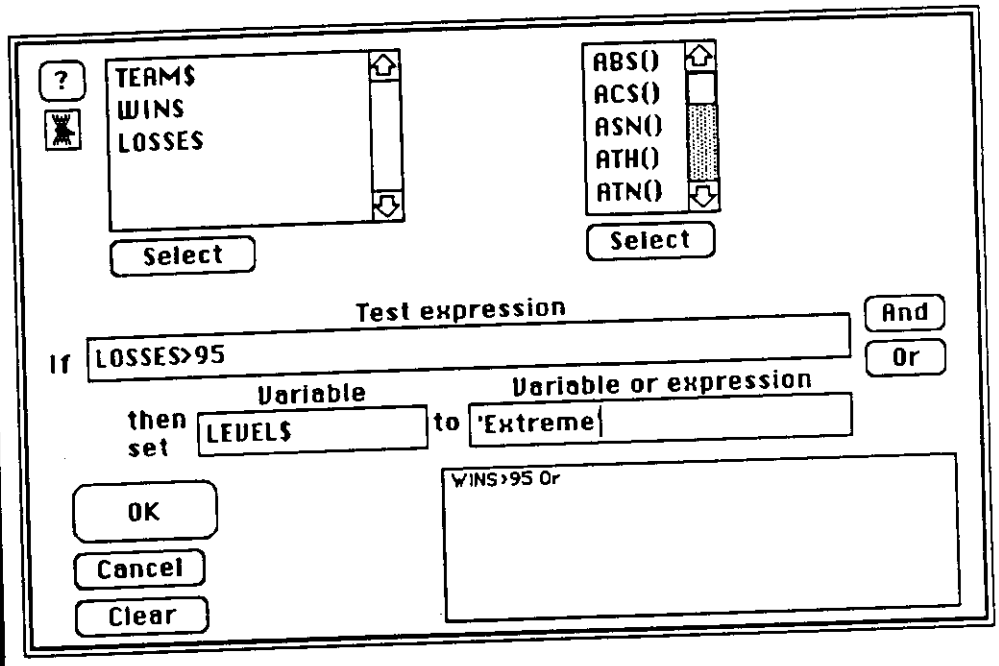


If you need to use [And] or [Or] in your test condition (your "if" clause), click [Complex recode...]. A new dialog box comes up, and we're ready to go.

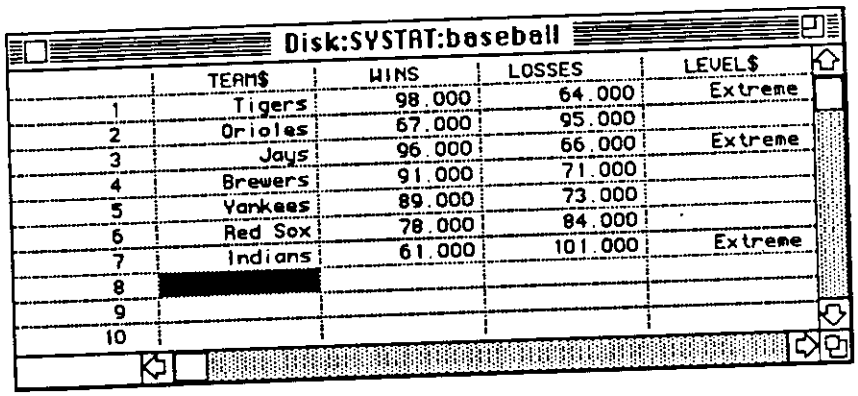
- Select **Recode...** from the **Editor** menu
- Click [Complex recode]
- Put together the test condition by selecting and typing:
  - Build the first condition: WINS>95
  - Click [Or]
  - Build the second condition: LOSSES>95
- Put together the "then" clause:
  - Type LEVEL$ in the second box
  - Type 'EXTREME' in the third box; don't forget the quotation marks
- Click OK

© 1989, SYSTAT, Inc.

Watch how the box in the lower right corner of the dialog box stores the parts of your test condition. You can keep clicking [And] and/or [Or] (try reading that ten times fast!) to add more conditions.



Again, we see the odometer and then a new variable appears.



	TEAM$	WINS	LOSSES	LEVEL$
1	Tigers	98.000	64.000	Extreme
2	Orioles	67.000	95.000	
3	Jays	96.000	66.000	Extreme
4	Brewers	91.000	71.000	
5	Yankees	89.000	73.000	
6	Red Sox	78.000	84.000	
7	Indians	61.000	101.000	Extreme
8				
9				
10				

Operators and functions

Following are the operators, relations, and functions available with **Math...** and **Recode...**

© 1989, SYSTAT, Inc.

Arithmetic operators

Arithmetic operators tell SYSTAT to perform a mathematical calculation between two numeric expressions.

+	addition
–	subtraction
*	multiplication
/	division
^	exponentiation

Here are examples of expressions you could use inside Set and If-then statements:

```
2*3
X-2
A/B
AGE^2
AGE^BEAUTY
```

SYSTAT sets the resulting values of inadmissible operations (e.g. division by zero) to missing.

Functions

Functions modify the variable, number, or expression you place inside the parentheses.

SQR	square root
LAG	lags variable by shifting values down one row
LOG	natural logarithm
EXP	exponential function
ABS	absolute value
SIN	sine
COS	cosine
TAN	tangent
ASN	arcsine
ACS	arccosine
ATN	arctangent
ATH	arc hyperbolic tangent (Fisher's *z*)
INT	integer truncation
LGM	log gamma
–	unary minus (negative)

© 1989, SYSTAT, Inc.

Here are examples of expressions you could use inside Set and If-then statements:

```
SQR(3)
INT(CARDIO+CANCER)
SIN(COS(TAN(3*Y)))
```

Arguments of trigonometric functions must be in radians, not degrees. LOG is the natural logarithm. You can obtain logs in other bases by dividing the natural log by the log of the base, e.g.:

```
LOG(X)/LOG(10)
LOG(INFO)/LOG(2)
```

SYSTAT sets the resulting values of inadmissible operations (e.g., square root of a negative number) to missing.

Relational operators

Relational operators are used in relational expressions to compare either two numeric or two character expressions.

<	less than
=	equal to
>	greater than
<>	not equal to
<=	less than or equal to
=>	greater than or equal to

Here are examples of expressions you could use inside Set and If-then statements:

```
AGE>BEAUTY
X <> .
GROUP >= 2
400<(CANCER+CARDIO)
STATE$='NY'
```

Logical operators

The logical operators [And] and [Or] connect two relational expressions to make a complex relational expression.

© 1989, SYSTAT, Inc.

Here are examples of expressions you could use inside Set and If-then statements:

```
AGE>17 And AGE<=60
SEX$='MALE' Or WEIGHT>300
(GROUP>5 Or SEX=1) And 400<(CARDIO+CANCER)
```

The logical operator [Not] reverses the value of an entire expression. [Not] sets any nonzero value (true) to zero, and any zero value (false) to one (true). For example:

```
If Not (PRICE<5000 And MAXSPEED>100 And PASSGR>4)
then set PURCHASE$ to 'FALSE'
```

Note that even if you are testing for different values on the same variable you must specify a separate relational expression for each value and join those expressions with logical operators. In the first example, AGE>17 And AGE<=60 is correct. The following, however, would *not* work:

```
AGE>17 Or <=60
60>=AGE>17
```

Some examples of legal statements using [And] and [Or] are:

```
If REGION=3 And CARDIO>420 then set RATE$ to 'HIGH'
Find CARDIO>430 Or CARDIO<400 And REGION$="WEST"
```

(Instructions for **Find...** appear in the "Data Editor utilities" section of this chapter.)

You can assign the results of a relational expression to a numeric variable. The variable receives a value of one if the expression is true; otherwise it receives a value of zero (false).

```
Set TRUTH to A > 0 Or B < 100
```

In this example, any case where A is greater than 0 or B is less than 100 will have value 1. Any case where A is less than 0 *and* B is greater than 100 will have value 0.

© 1989, SYSTAT, Inc.

(Note: SYSTAT follows the standard for programming languages and returns 0 for false and 1 for true. Microsoft Basic is nonstandard and returns 0 for false and –1 for true.)

You may not use logical operators to join two "then" statements. The following statement using And and Or is not possible:

```
If GROUP=1 then set TEST to 1 And set TRIAL to 2
```

Multi-variable functions

Cross-variable functions, unlike other SYGRAPH operators, take only variables as arguments. They are also unique in that they use only nonmissing values in their calculations and exclude any variables that are missing.

AVG	mean of nonmissing values
SUM	sum of nonmissing values
MIN	minimum value of values
MAX	maximum value of values
STD	standard deviation of nonmissing values
MIS	number of missing values

If we wanted to find the mean of the three variables A, B, and C for each case, we could use either of the statements:

```
Set MEAN to (A+B+C)/3
Set MEAN to AVG(A,B,C)
```

In the first statement, MEAN is set to missing when any of the three variables is missing. In the second statement, AVG returns the mean of the nonmissing variables. MEAN is missing only if all three are missing.

You can use subscripts if you want to sum the values in an array or set of contiguous variables:

```
Set TOTAL to SUM(X(1-20))
Set TOTAL to SUM(A-Z)
```


© 1989, SYSTAT, Inc.

Distribution functions

There are three kinds of distribution functions: cumulative, inverse, and random variate.

Distribution	*Cumulative*	*Inverse*	*Random data*
Uniform	UCF(*x*)	UIF(a)	URN
Normal	ZCF(*z*)	ZIF(α)	ZRN
t	TCF(*t*,*df*)	TIF(α,*df*)	TRN(*df*)
F	FCF(*F*,*df1*,*df2*)	FIF(α,*df1*,*df2*)	FRN(*df1*,*df2*)
Chi-square	XCF(χ^2,*df*)	XIF(α,*df*)	XRN(*df*)
Exponential	ECF(*x*)	EIF(α)	ERN
Gamma	GCF(γ,*p*)	GIF(α,*p*)	GRN(*p*)
Beta	BCF(β,*p*,*q*)	BIF(α,*p*,*q*)	BRN(*p*,*q*)

Cumulative functions compute the probability that a random value from the specified distribution falls below the given value (i.e. the proportion of the distribution below that value). *Inverse functions* take a specified α (a probability value between zero and one) and return the critical value below which lies that proportion of the specified distribution. *Random variate functions* generate pseudo-random variates from the specified distribution.

The algorithms for all functions were chosen to favor numerical accuracy over speed. Some functions may take time to evaluate for some choices of parameters. Please be patient.

SYGRAPH supplies eight distributions: uniform, normal, *t*, *F*, chi-square, exponential, gamma, and beta. Uniform distribution refers to a distribution uniformly distributed from zero to one. Normal distribution is the standard normal distribution N~(0,1) with mean zero and standard deviation one. Exponential has parameter one. All the others require you to specify their parameters. These parameters may be non-integer.

Use the cumulative distribution functions to obtain probabilities associated with observed sample statistics. For many applications, to obtain "significance levels" from the FCF or XCF functions you must find the upper tail probabilities, e.g.:

```
Set P to 1 - FCF(4.14,3,7)
```

© 1989, SYSTAT, Inc.

You can use the functions we provide to evaluate probabilities for distributions for which there are no explicit cumulative distribution functions. For example, you can evaluate the cumulative binomial probability of *k* or more occurrences in *n* trials with binomial probability *p* using the BCF function:

```
Set BINPROB to BCF(P,K,N-K+1 )
```

Using the GCF function, you can evaluate the cumulative Poisson probability of the number of random events between zero and *k* when the expected value is *X*:

```
Set POISSONP to 1 - GCF(X,K+1)
```

Use the inverse distributions to determine critical values and to construct confidence intervals. They are also handy for power calculations.

You can use transformations of the random variate functions to produce random deviates of the desired forms. To produce a random normal deviate with the required mean and standard deviation use one like:

```
Set NORMAL to MEAN+STDDEV*ZRN
```

To produce a uniform deviate between A and B use an expression like:

```
Set UNIFRMAB to A+(B-A)*URN
```

To get an exponential variate with parameter lambda, use:

```
Set EXP to ERN/LAMBDA
```

To get an extreme value variate with location alpha and scale beta, use:

```
Set EXTREME to ALPHA - BETA*LOG(ERN)
```

To get a gamma variate with shape alpha and scale beta, use:

```
Set GAMMA to BETA*GRN(ALPHA)
```

If ALPHA is an integer, then the above expression returns an Erlang variate.

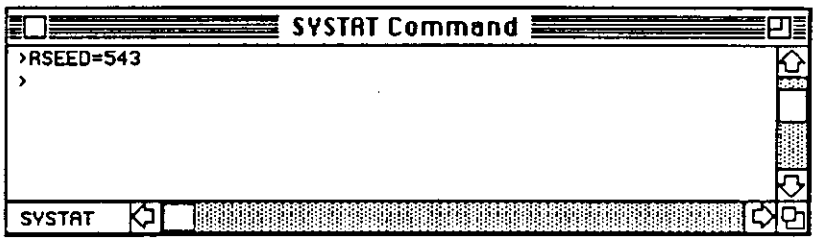
© 1989, SYSTAT, Inc.

Random number generators

The algorithm for uniform variates generates the numbers by a triple modulo method. Each uniform is constructed from three multiplicative congruential generators with prime modulus. The initial seeds for each generator are 13579, 12345 and 313 (Wichman and Hill, 1982). You can modify the last seed of this generator by using the RSEED command. Select **Show command window...** from the **Goodies** menu and enter the command:

```
RSEED=seed
```

where *seed* is an integer between 1 and 30,000. Press Return to execute the command.



The pseudo-random variates for the other distribution functions are generated by applying the appropriate inverse cumulative distribution function to uniform random variates between zero and one.

Note: if you have been using SYSTAT Version 3.1 or earlier, note that the names of the SYSTAT distribution functions have been changed.

Old name	*New name*
URAN	URN
NRAN	NRN
CDF	ZCF
IDF	ZIF

© 1989, SYSTAT, Inc.

Order of expression evaluation

Expressions are evaluated from left to right according to the precedence of operators. That is, operators with higher precedence are evaluated before those with lower. Order of precedence from highest to lowest runs as follows:

1) Expressions enclosed in parentheses	( )
2) Exponentiation	^
3) Unary minus	–
4) Multiplication and division	*, /
5) Addition and subtraction	+, -
6) Relational operators	=, <>, <, >, <=, >=
7) Logical operators	**And, Or**
8) Logical negation	**Not**

This precedence is the same as in the BASIC programming language (which is discussed in *DATA*). Because of this order of precedence, the expression

A+B*C

is the same as

A+(B*C)

and not the same as

(A+B)*C

Multiplication (*) has a higher precedence than addition (+) and is therefore performed first. If A=1, B=2, and C=3, the above expressions evaluate as follows:

A+B*C = 1+2·3 = 1+6 = 7
A+(B*C) = 1+(2·3) = 1+6 = 7
(A+B)*C = (1+2)·3 = 3·3 = 9

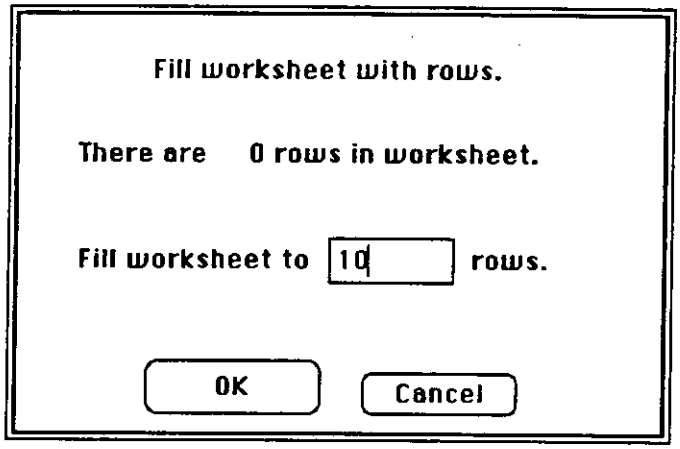
Fill worksheet...

Fill worksheet... from the **Editor** menu fills a specified number of cases with missing values.

© 1989, SYSTAT, Inc.

You must first name the variables you want to fill.

- Select **Fill worksheet...** from the **Editor** menu
- Specify the number of cases to fill
- Click OK



Filling variables with random data

Once you've filled a variable or set of variables with missing values, you can use **Math...** to fill the variables with random values using the URN, ZRN, TRN, FRN, XIF, ERN, GRN, and BRN functions discussed above.

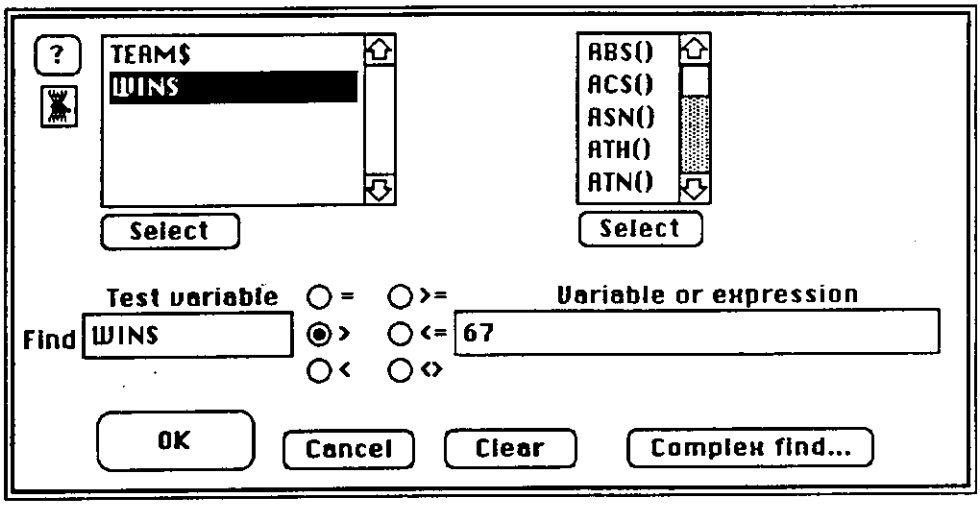
For example, here we fill the variable A with normal random data.

- Select **Math...** from the **Editor** menu
- Fill in the statement: Set A to ZRN
- Click OK

© 1989, SYSTAT, Inc.

Find case...

You can search your data for values meeting specific conditions with **Find case...** from the **Editor** menu.



You can use the functions, operators, and relations discussed above for your Find statement. If you need to use the logical operators [And], [Or], or [Not], click [Complex find].

Find... starts its search from the current case (wherever the cursor is resting) and goes down the file. When SYSTAT finds a case that meets your condition, it moves the cursor to that case in the (first) variable named in your Find statement.

Find next

After a **Find...** is done, **Find next** becomes available in the **Editor** menu. **Find next** executes a new search on the same Find statement.

Try using **Find case...** on the USDATA dataset.

- Select **Open...** from the **File** menu
- Select the USDATA dataset from the Data Files folder
- Click [Edit]
- Click Open

© 1989, SYSTAT, Inc.

The cursor should be resting in the top row. Try searching for Michigan.

- Select **Find case...** from the **Editor** menu
- Fill in the statement: Find STATE$ = 'MI'
- Click OK

The cursor moves to the STATE$ column on case 13.

Disk:SYSTAT:USDATA

	STATE$	REGION$	REGION	DIVISION$
13	MI	Midwest	2.000	E N Central
14	WI	Midwest	2.000	E N Central
15	MN	Midwest	2.000	W N Central
16	IA	Midwest	2.000	W N Central
17	MO	Midwest	2.000	W N Central
18	ND	Midwest	2.000	W N Central
19	SD	Midwest	2.000	W N Central
20	NB	Midwest	2.000	W N Central
21	KS	Midwest	2.000	W N Central
22	DE	South	3.000	S Atlantic

Now, move the cursor back to the top of the file and do another search.

- Select **Find case...** from the **Editor** menu
- Fill in the statement: Find REGION = 2
- Click OK

Disk:SYSTAT:USDATA

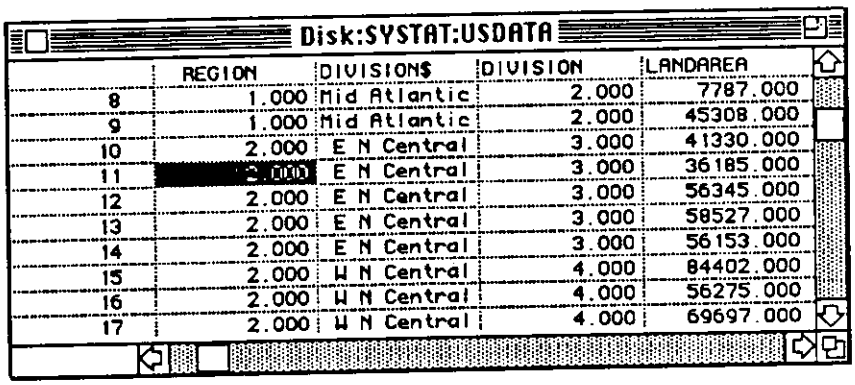
	STATE$	REGION$	REGION	DIVISION$
1	ME	Northeast	1.000	New England
2	NH	Northeast	1.000	New England
3	VT	Northeast	1.000	New England
4	MA	Northeast	1.000	New England
5	RI	Northeast	1.000	New England
6	CT	Northeast	1.000	New England
7	NY	Northeast	1.000	Mid Atlantic
8	NJ	Northeast	1.000	Mid Atlantic
9	PA	Northeast	1.000	Mid Atlantic
10	OH	Midwest	2.000	E N Central

Now, repeat the search.

- Select **Find next** from the **Editor** menu

The next case with REGION=2 is case 11.

© 1989, SYSTAT, Inc.

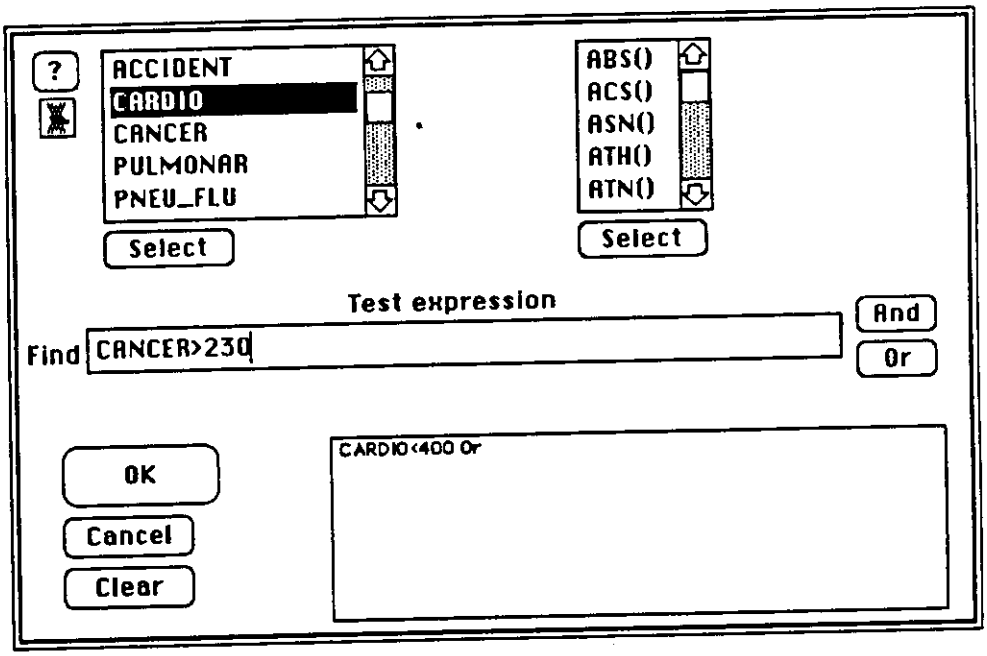


	REGION	DIVISION$	DIVISION	LANDAREA
8	1.000	Mid Atlantic	2.000	7787.000
9	1.000	Mid Atlantic	2.000	45308.000
10	2.000	E N Central	3.000	41330.000
11	2.000	E N Central	3.000	36185.000
12	2.000	E N Central	3.000	56345.000
13	2.000	E N Central	3.000	58527.000
14	2.000	E N Central	3.000	56153.000
15	2.000	W N Central	4.000	84402.000
16	2.000	W N Central	4.000	56275.000
17	2.000	W N Central	4.000	69697.000

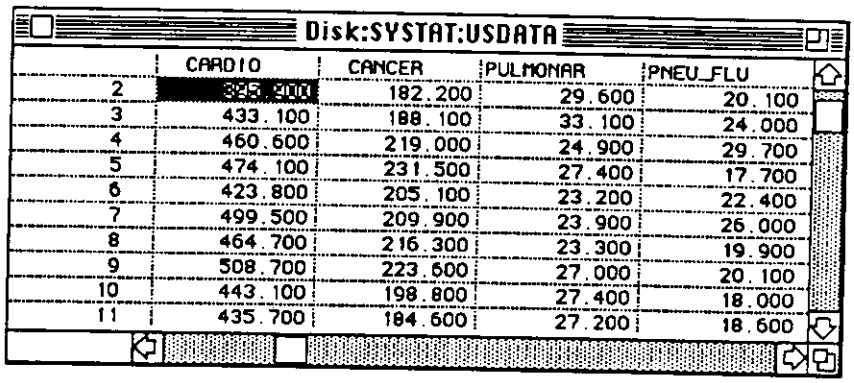
Now, let's try a complex search. Move the cursor back to the top of the file.

- Select **Find case...** from the **Editor** menu
- Click [Complex find]
- Build the statement: Find CARDIO<400 [Or] CANCER>230

The moves to the second case in the CARDIO column, the first case where CARDIO is less than 400 or CANCER is greater than 230. In this instance, the first expression in the statement, CARDIO<400, is satisfied.



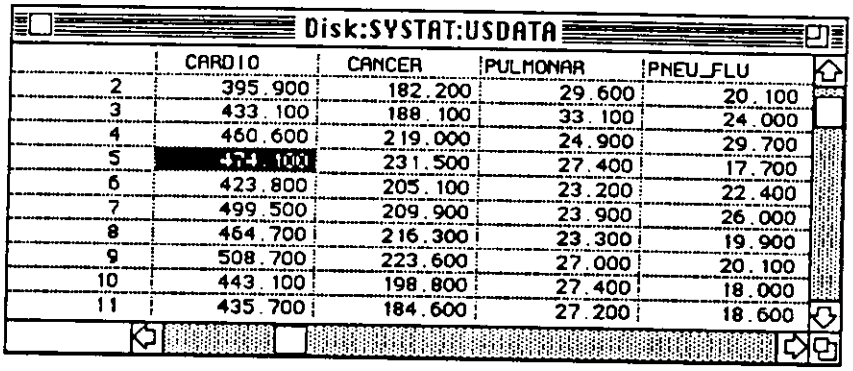
© 1989, SYSTAT, Inc.



Disk:SYSTAT:USDATA

	CARDIO	CANCER	PULMONAR	PNEU_FLU
2	395.900	182.200	29.600	20.100
3	433.100	188.100	33.100	24.000
4	460.600	219.000	24.900	29.700
5	474.100	231.500	27.400	17.700
6	423.800	205.100	23.200	22.400
7	499.500	209.900	23.900	26.000
8	464.700	216.300	23.300	19.900
9	508.700	223.600	27.000	20.100
10	443.100	198.800	27.400	18.000
11	435.700	184.600	27.200	18.600

Now try **Find next**, which sends the cursor to the fifth case.



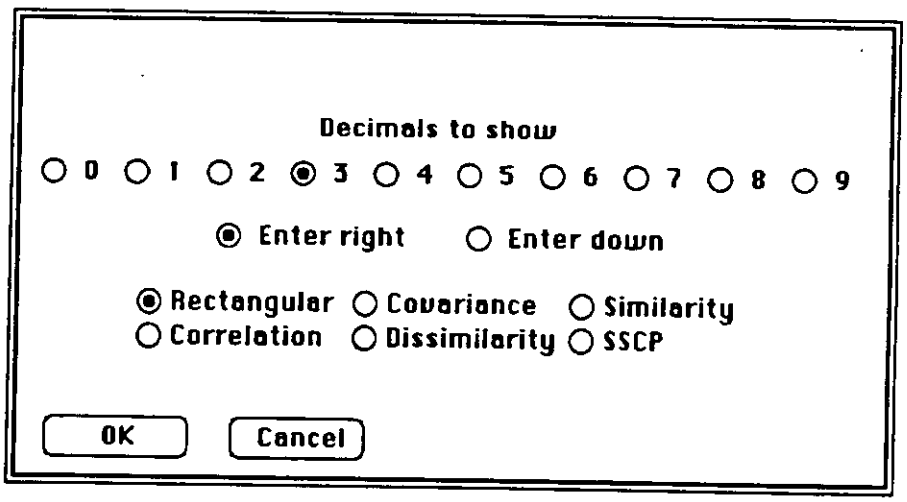
Disk:SYSTAT:USDATA

	CARDIO	CANCER	PULMONAR	PNEU_FLU
2	395.900	182.200	29.600	20.100
3	433.100	188.100	33.100	24.000
4	460.600	219.000	24.900	29.700
5	474.100	231.500	27.400	17.700
6	423.800	205.100	23.200	22.400
7	499.500	209.900	23.900	26.000
8	464.700	216.300	23.300	19.900
9	508.700	223.600	27.000	20.100
10	443.100	198.800	27.400	18.000
11	435.700	184.600	27.200	18.600

Note that even though the case meets the condition CANCER>230, the cursor is in the CARDIO column, since CARDIO is the first variable in the Find statement.

Editor/ Formats...

Formats... in the **Editor** menu lets you control characteristics of the Data Editor.



© 1989, SYSTAT, Inc.

Decimals to show

In the Data Editor, numbers are displayed with three decimal places by default. If you want to change this, choose a different number.

Your choices affect only the current display and not the way the data are stored; the next time you open the file, the numbers will have three decimal places. If you always want to see the same number of decimals, you can record your choice with **Record preferences** from the **Goodies** menu.

Enter right/down

After you enter a data point by pressing Return, the cursor jumps one cell to the right. If you want it to jump one cell down instead, choose [Enter down].

Matrix type

Data files are [Rectangular] by default. If you are entering a matrix (see triangular matrices below), choose the appropriate type. If you have opened a matrix file generated by SYSTAT, the correct type will be shown. Thus, you can use **Formats...** to find out what type file you have.

Help

You can get help for the Data Editor with **Help** from the **Editor** menu. This and other help items are discussed in the previous chapter, **Help**.

Select...

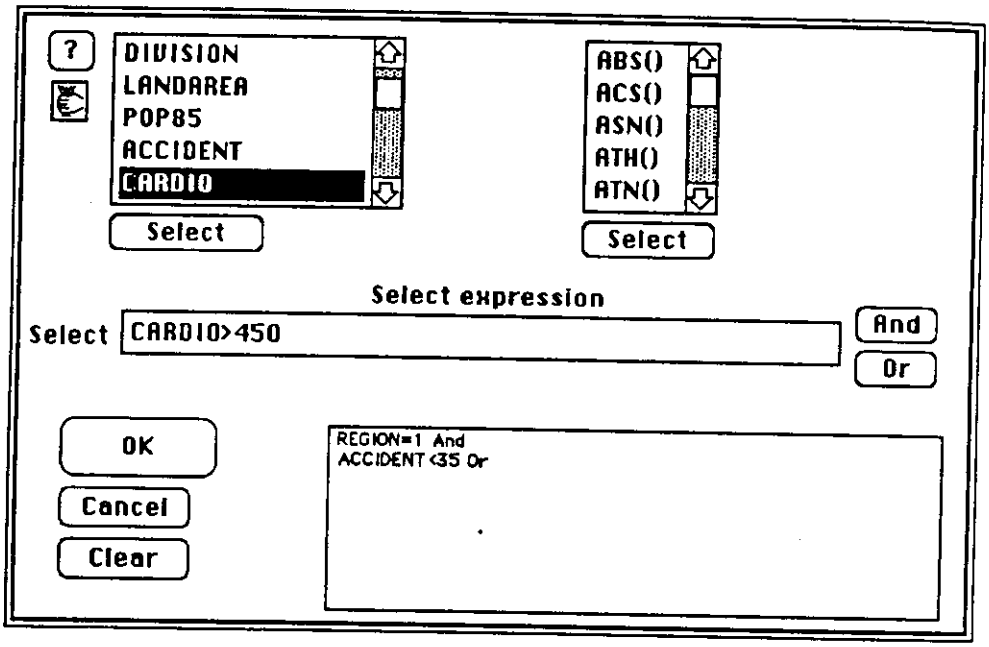
Select... from the **Editor** menu selects subgroups of cases according to the test condition(s) you specify. Subsequent analyses use only the cases you select.

You can link several test conditions together with the logical [And] and [Or] buttons. The **Select...** dialog box provides a selection list of variables and another of SYSTAT's functions for your convenience; you may build the Select statement with any combination of selecting, typing, and using the [And] and [Or] buttons.

- Select **Select...** from the **Editor** menu
- Build the Select statement; click [And] or [Or] between expressions as necessary
- Click OK

© 1989, SYSTAT, Inc.

```
Select REGION=1 And ACCIDENT<35 Or CARDIO>450
```



Oops! If you try this, you'll discover that fourteen cases are found, and some of them do not have REGION=1. This is because SYSTAT reads Select statements from left to right, so the implicit grouping is:

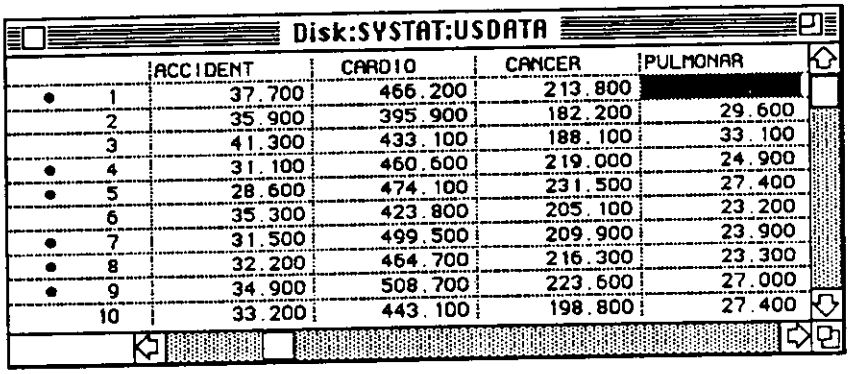
```
Select (REGION=1 And ACCIDENT<35) Or CARDIO>450
```

With our Select statement, any case that is in region 1 and ACCIDENT<35 is selected, and any case *in any region with any accident rate* that has CARDIO>450 is also selected. What we want is actually the following:

```
Select REGION=1 And (ACCIDENT<35 Or CARDIO>450)
```

Now we get what we want. SYSTAT reports "6 cases were found" and marks them with bullets in the CASE column of the Editor. We scrolled right in the Editor window so you could see the ACCIDENT and CARDIO values. All six cases are in Region 1.

© 1989, SYSTAT, Inc.



Disk:SYSTAT:USDATA

	ACCIDENT	CARDIO	CANCER	PULMONAR
1	37.700	466.200	213.800	
2	35.900	395.900	182.200	29.600
3	41.300	433.100	188.100	33.100
4	31.100	460.600	219.000	24.900
5	28.600	474.100	231.500	27.400
6	35.300	423.800	205.100	23.200
7	31.500	499.500	209.900	23.900
8	32.200	464.700	216.300	23.300
9	34.900	508.700	223.600	27.000
10	33.200	443.100	198.800	27.400

For graphics and analyses only selected cases are used. If we were to use **Corr** now, for example, we would get a correlation matrix based only on those six cases. However, if you use **Math...**, **Recode...**, or **Select...** the entire data set is used.

Random subsamples

You can select random subsamples of data with **Select...** For example,

```
Select URN<.5
```

selects a random subset of approximately 50% of the cases.

Extract selected...

You can do analyses on selected data either by proceeding from here with cases bulleted in the Editor, or by extracting the data to their own data file and using that file. It is handy to extract cases to a file if you anticipate needing to use them in future SYSTAT sessions. Otherwise, the temporary selection is sufficient.

To extract selected cases and store them in a separate file, use **Extract selected...** in the **Editor** menu. (You must select cases as shown above before you can extract them.)

- Select **Extract selected...** from the **Editor** menu
- Specify a filename and folder location
- Click Save

You can open the new data file to analyze the extracted cases.

Remove select

Select... stays in effect until you quit SYSTAT, close the Editor, or open a new data file. To turn off **Select...**, select **Remove select** from the **Editor** menu.

© 1989, SYSTAT, Inc.

Other ways to select data

The **Select...** item in the **Data** menu also lets you do case selection for analyses. **Data/Select...** is discussed in the **Analysis** chapter later in this volume.

You can also do case selection with the scatterplot brushing tools discussed in the **Exploratory data analysis** chapter.

Here is a quick comparison of the three ways of doing selection:

	Editor/Select...	Brushing tools	Data/Select...
Method of selecting cases	algebraic	visual	algebraic
Logical operators available	And, Or	—	And
Selected cases marked	in Editor	in Editor and plot	—
Windows that must be open	Editor	Editor and View	—

The Editor and brushing methods allow cumulative selection (selecting more cases to add to the selection) and all three allow quick de-selection. Obviously, the scatterplot brushing tools only work with scatterplots

In general, **Editor/Select...** is more powerful and brushing tools are more intuitive than **Data/Select...** However, **Data/Select...** is more convenient when it is preferable not to have the Data Editor open, since having the file available in the Editor can slow SYSTAT down.

© 1989, SYSTAT, Inc.

The datasets we've looked at so far have been fairly conventional: one variable for Case (automatically provided by SYSTAT) and individual variables recording certain characteristics for each case. There are a few different types of data formats you might need: grouping variables, aggregated data, and triangular matrices.

Grouping variables

A task familiar to statisticians is determining whether categories of cases differ from each other. You might want to compare weight losses between men and women, or popularity between colas and non-colas, prices between different types of stocks, and so on.

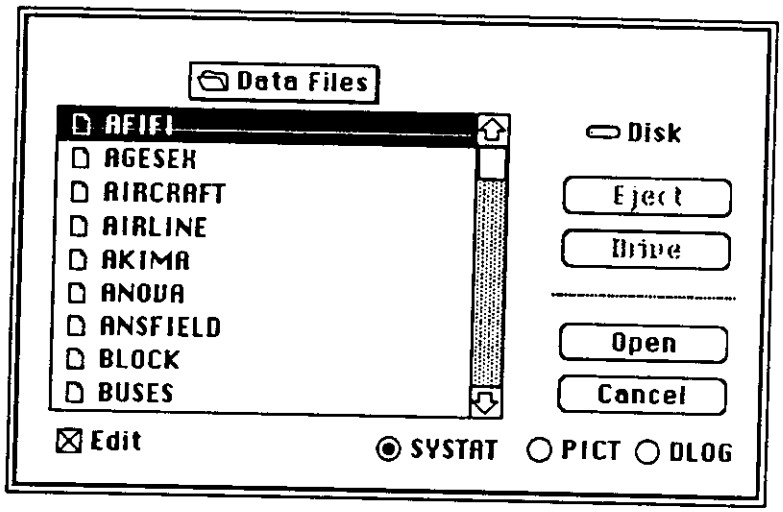
You can do this by using *grouping variables*. A grouping variable is a numeric or character variable that does nothing more than label each case with its group.

Grouped datasets let you do comparison and contrast tests: grouped box plots, ANOVAs, chi-square tests, summary statistics by groups, category plots and bar charts, etc.

The AFIFI dataset records increases in systolic blood pressure after administering one of four different drugs to patients with one of three different diseases. Open the AFIFI dataset and take a look at it.

- Select **Open...** from the **File** menu
- Click [Edit]
- Choose AFIFI from the Data Files folder
- Click Open

© 1989, SYSTAT, Inc.



DRUG and DISEASE are both grouping variables. Scroll through the dataset and notice how DRUG has only four values, 1, 2, 3, and 4. DISEASE has three values, 1, 2, and 3. Each patient in the study is a separate row, or case. You can see which disease each one had by looking at the value for DISEASE in the row, and similarly which DRUG the patient was given.

Disk:SYSTAT:AFIF1

	DRUG	DISEASE	SYSINCR
1	1.000	1.000	42.000
2	1.000	1.000	44.000
3	1.000	1.000	36.000
4	1.000	1.000	13.000
5	1.000	1.000	19.000
6	1.000	1.000	22.000
7	1.000	2.000	33.000
8	1.000	2.000	26.000
9	1.000	2.000	33.000
10	1.000	2.000	21.000

We see, for example, that the seventh patient (the row where Case is 7) has the DISEASE 2 and was given DRUG 1. After the drug was administered, Mr. Seven (Ms.Seven) had an increase of 33 in his (her?) systolic blood pressure.

We will see this dataset used in the the **MGLH** chapter of the SYSTAT manual for two-way analysis of variance.

© 1989, SYSTAT, Inc.

Aggregated data

Aggregated datasets group objects with the same measurement values as a single case in the file. The CHEST dataset counts the number of soldiers with given chest measurements. With the regular data format, we would enter a separate case for each soldier, with variables like NAME$, CHEST_IN, AGE, etc. This format is great if you want to record several characteristics, but we'd spend days typing Mr. Quetelet's 5738 cases. This would be silly, especially since we don't care about any other characteristics. A frequency distribution would be more manageable and more useful.

Here is how the data look in a frequency distribution. The unique values in this study are the chest measurements. So, we name a variable CHEST_IN, and list all the different measurements down the column. Then, we add a count variable, COUNT.

CHEST_IN	COUNT
33	3
34	18
35	81
36	185
37	420
38	749
39	1073
40	1079
41	934
42	658
43	370
44	92
45	50
46	21
47	4
48	1

As you can see, this is a sensible way to set up the data. Interestingly, we can even read the COUNT values as a rough histogram: note that the numerals themselves (as well as their values) widen in the middle and taper off at the ends.

The DANISH dataset combines grouping variables with a COUNT variable. This dataset is provided in the Data Files folder.

© 1989, SYSTAT, Inc.

FATHER	SON	COUNT
1	1	18
1	2	17
1	3	16
1	4	4
1	5	2
2	1	24
2	2	105
2	3	109
2	4	59
2	5	21
3	1	23
3	2	84
3	3	289
3	4	217
3	5	95
4	1	8
4	2	49
4	3	175
4	4	348
4	5	198
5	1	6
5	2	8
5	3	69
5	4	201
5	5	246

Both FATHER and SON are grouping variables indicating to which economic class the father or son belonged. The COUNT variable shows how many combinations existed. We can see, for example, that of the 2391 father-son pairs included in this study, 18 pairs had both father and son from the 1st class. 217 pairs involved class 3 fathers who had class 4 sons. As you can see, relatively few sons were able to rise to an economic class above that of their fathers.

We analyze the DANISH dataset in the **Tables** chapter of the SYSTAT manual to see what patterns exist among fathers' and sons' economic classes.

© 1989, SYSTAT, Inc.

Triangular matrices

Sometimes you may wish to input your data in triangular form. Corr, for example, computes correlation, similarity, and dissimilarity matrices. A correlation matrix shows the correlation coefficients for each pairing of variables:

	ACCIDENT	CARDIO	CANCER	PULMONAR	PNEU_FLU
ACCIDENT	1.000				
CARDIO	-0.506	1.000			
CANCER	-0.578	0.908	1.000		
PULMONAR	-0.099	0.441	0.438	1.000	
PNEU_FLU	-0.196	0.538	0.358	0.400	1.000

This matrix is included in the **Corr/Pearson...** output for five variables from the MEDICAL dataset. To find out the correlation between ACCIDENT and CANCER, for instance, look at the cell in the CANCER row and the ACCIDENT column. The correlation is –0.578. Note that correlation matrices are symmetric, so only the lower diagonal portion is shown. Since variables always correlate perfectly with themselves, the diagonal always has trivial values of 1.

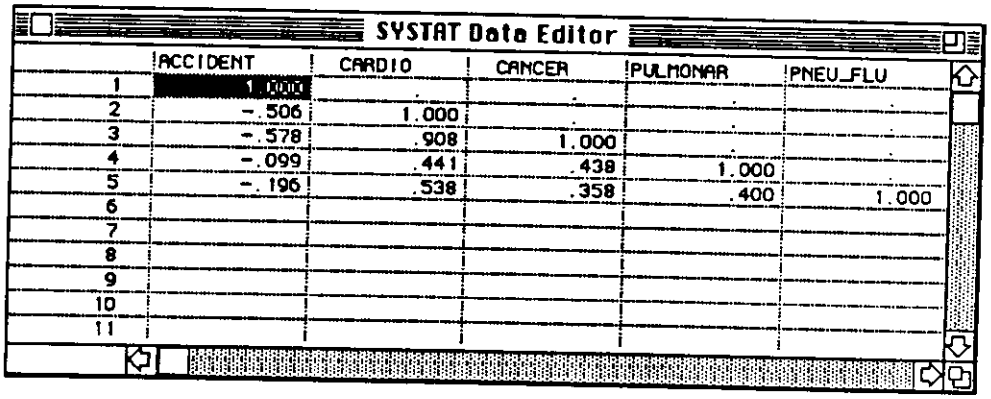
Suppose you wanted to factor this correlation matrix. You would just save a file with **Corr**, open the file, and factor it with **Factor**. The upper diagonal portion is filled with missing values, so the file would look something like this:

ACCIDENT	CARDIO	CANCER	PULMONAR	PNEU_FLU
1.000	.	.	.	.
-0.506	1.000	.	.	.
-0.578	0.908	1.000	.	.
-0.099	0.441	0.438	1.000	.
-0.196	0.538	0.358	0.400	1.000

Now, suppose you wanted to factor (or do regression or statistics or graphs...) a matrix that somebody else published and you don't have the original data. Or perhaps you want to type in a dissimilarity matrix directly for **Cluster** analysis.

The Editor lets you input matrices directly. As usual, variable names go in the top row. Don't label the cases like the correlation output does; the file should look like the matrix directly above.

© 1989, SYSTAT, Inc.



After you enter the data, you must tell SYSTAT what type of matrix it is with **Formats...** from the **Editor** menu. It is important to do this *before* you save the file.

- Select **Formats...** from the **Editor** menu
- Choose the type of matrix: [Correlation]
- Click OK

As you can see, the default is [Rectangular]. Rectangular files are the usual files of variables and cases. The other five types are triangular matrices that can be produced with **Corr.** [SSCP] stands for Sums of Squares and Cross Products. (See the **Corr** chapter in the *SYSTAT* manual for more information on these matrix types.)

© 1989, SYSTAT, Inc.

5 Data files

Saving data in files 94
Filenames 94
Saving data 94
Save... 94
Save as... 95
Single and double precision 96
Problems? 97

Printing data 98
Page Setup... 98
Print data... 98
Other ways to print data 98
Problems? 98

Importing data 99
Pasting data from the Clipboard 99
Reading text files 100
Reading Excel files 102
Problems? 103
Reading map files 103

Exporting data 104
Copying or Cutting into the Clipboard 104
Saving text files 104

Manipulating data files 106
Sort... 106
Rank... 108
Standardize... 109
Transpose... 109

© 1989, SYSTAT, Inc.

Data files 5

Overview

This chapter discusses data files: how to name them, save them, print them, exchange them with other applications, sort them, rank them, standardize them, and transpose them.

We discuss saving and printing *output* (data files or text produced by statistical analyses) in the next chapter, **Redirecting output**.

© 1989, SYSTAT, Inc.

Filenames

SYSTAT filenames are subject only to the usual restrictions imposed by Macintosh conventions. You may use upper- or lower-case characters, spaces, and some symbols. If you try to specify an illegal filename, you will receive an error message.

The data files that we provide with SYSTAT in the Data Files folder are named in all upper-case characters. Again, this is not necessary. We have just chosen to name files this way so you can always recognize filenames easily when reading our text.

Previous versions of SYSTAT imposed some additional limitations on filenames; these limitations are no longer in force. You no longer are limited to eight upper-case alphanumeric characters: you may now use periods, underscores, spaces, and numbers. You no longer need to use the .SYS, .CMD, and .DAT filename extensions (suffixes). If you have files created with previous versions, the old filenames will work fine with this version. You do not need to rename any old files unless you want to do so.

Note that if you choose to use the optional command interface, you must surround folders and filenames with single or double quotation marks.

Also note that SYGRAPH map files must still have a .MAP suffix. We discuss .MAP files in more detail in an appendix to the *SYGRAPH* volume.

Saving data

You can save data with two items of the **File** menu. **Save** saves your data (or your changes) to the current file, overwriting the previously-stored data. **Save as...** provides a filename which you can change. **Save as...** also lets you save your data in different formats suitable for exporting to another application.

Save

Save from the **File** menu saves your data in the current file and location (folder), writing over any previous data that may have been stored under that filename.

© 1989, SYSTAT, Inc.

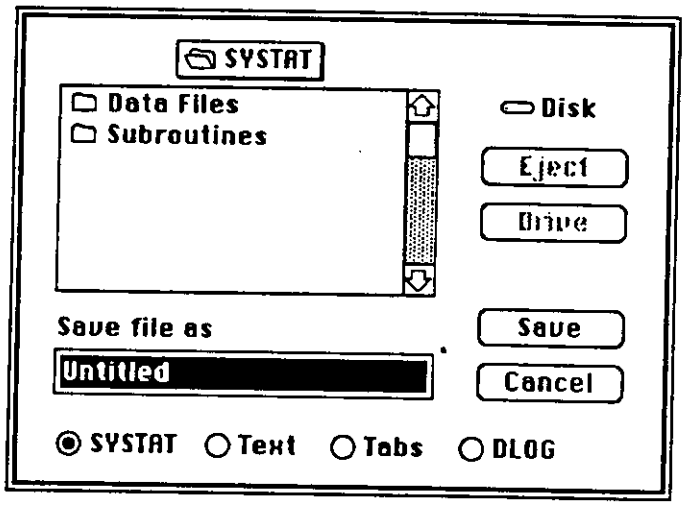
If you started with a **New** worksheet, **Save** gives you the Save as... dialog box (discussed next) so you can specify a filename and folder location.

- Select **Save** from the **File** menu
- Specify a filename and location
- Click Save

Save as...

Save as... from the **File** menu lets you specify a filename and folder location. You may also save your data files in one of two plain text (ASCII) formats, [Text] and [Tabs.]

- Select **Save as...** from the **File** menu
- Specify a filename and folder location
- Choose a text format for exporting
- Click Save



SYSTAT format

The default file format is a SYSTAT file. If you are going to do statistical or graphical analyses on the data you save, they must be in SYSTAT format.

Plain text formats

You can save SYSTAT data files in one of two plain text formats. [Text] saves your data in comma-delimited form, where successive values in a row are separated by commas and lines are separated with carriage returns. [Tabs] saves in tab-delimited form, where successive values in a row are separated by tabs into columns; lines are separated with carriage returns.

© 1989, SYSTAT, Inc.

You may read plain text data files into other applications such as Microsoft Excel. Then, you can save results from that application in text form and read them back into SYSTAT with **Open...** from the **File** menu.

Single and double precision

SYSTAT saves data files from the editor in double precision by default. Double precision allots 8 bytes per value, which retains up to 15 significant digits in your data.

You can also save in single precision, which allots 4 bytes per value and ensures approximately 7 significant digits. Saving in single precision takes up about half the disk space as double. SYSTAT does all its calculations in double precision, so it converts single precision data to double before processing.

All character values are stored in 12 byte strings regardless of whether you save in single or double precision.

Again, SYSTAT saves data from the editor in double precision by default. To save data in single precision, you must use the command interface. Select **Show command window** from the **Goodies** menu. Then, type at the prompt, pressing Return at the end of the command.

```
>WRITE filename / SINGLE
```

In the command above, *filename* represents the filename you give to the file. If you want to save in a different folder or specify a filename that includes blank spaces, you must surround the complete folder location and filename with single or double quotation marks:

```
>WRITE 'Disk:SYSTAT:My Data:myfile' / SINGLE
```

For more information about the optional command interface, see the **Command reference** chapter at the back of each manual.

When you save data produced by statistical procedures, a dialog box asks whether to save in single or double precision

© 1989, SYSTAT, Inc.

Problems?

Here are some guidelines for avoiding problems.

1) Be sure you have enough space on your hard disk. As a general rule, you should try to make sure that you have about four times more free disk space than your largest data file. (You can see how much space you have by looking at the window of your topmost "folder"—the one you see by double-clicking the hard disk icon. Select **By icon** from the **View** menu befored starting SYSTAT. The top right corner of the window shows how much hard disk space is free.)

2) Save frequently whenever you are working with data files. That way, if the power fails or you spill tea on your machine, you will only lose a small amount of work.

3) Because SYSTAT operates both on commands and menus, it is a good idea not to include apostrophes in folder or filenames. SYSTAT, like most other computer languages, uses apostrophes as character delimiters in commands. While it is trivial to process apostrophes through menus, there are conflicts you can encounter when they appear in commands. If you have problems, such as an "empty file" error, just remove the apostrophe and try again.

© 1989, SYSTAT, Inc.

Printing data

You can use **Print data...** from the **Editor** menu to print your data. **Page Setup** in the **File** menu lets you set up how things are to be printed from SYSTAT.

Page Setup...

Page Setup... lets you choose options for printing. Each type of printing device has its own dialog box, so what you see in the dialog box depends on what kind of printing device (LaserWriter, ImageWriter, etc.) you have hooked up and what version of the printer drivers you are using.

Print data...

Print data... prints the data in the file currently in the Data Editor window.

- Select **Print data...** from the **Editor** menu

Other ways to print data

You can use text files or the Clipboard and **Cut**, **Copy**, and **Paste** to move all or part of a worksheet into a text editing application (such as WordPerfect or Microsoft Word) where you can format the data to your needs before printing. Instructions for exporting data to other applications through text files or the Clipboard appear in "Exporting" section of this chapter.

You can use the DATA facilities to print all or part of a dataset. See the first several chapters of the *DATA* volume for instructions.

Problems?

If you have problems printing data or output from SYSTAT, be sure that you have current versions of the appropriate drivers (for a LaserWriter, for example, you need both LaserWriter and LaserPrep) present in your System Folder. Often problems occur because these files are either missing or damaged, and you can fix things by merely recopying the files from the master disk, restarting your Macintosh, and trying again.

© 1989, SYSTAT, Inc.

You can import data from other applications three different ways.

1) You can **Cut** or **Copy** data in the other application, switch to SYSTAT, and **Paste** the data into the SYSTAT Data Editor.

2) You can save data from one application in a text file, switch to SYSTAT, and read the text file into the Data Editor.

3) You can use SYSTAT's **Import...** facility to read Excel data files or to do fast importing of ASCII text files.

The Clipboard technique is a quick way to import small amounts of data. For larger files, we recommend using external text files. Also, you need to use text files when importing data from other platforms, such as from a UNIX machine.

Pasting data from the Clipboard

You can transfer data from another application to SYSTAT through the Clipboard.

You can paste into a new worksheet or an existing data file.

- Open the other application
- Select the range of information
- Select **Cut** or **Copy** from the **Edit** menu
- Quit from the application

- Start SYSTAT
- Select **New** from the **File** menu and label the variables, or **Open...** an existing SYSTAT file
- Click the upper right corner cell of the range where you want the data to appear
- Select **Paste** from the **Edit** menu
- Save your new data

Delimiters

The pasted data must be delimited by commas, spaces, or tabs.

© 1989, SYSTAT, Inc.

Variable names

If you do not transfer variable names with your data, you must label the variable columns before pasting the data. If you don't, SYSTAT names the columns COL01, COL02, ..., COL*n*, and of course you can edit those names after you paste the data.

Character values

Any character values that contain spaces or single or double quotation marks must be surrounded with (the opposite type of) quotation marks.

Reading a text file

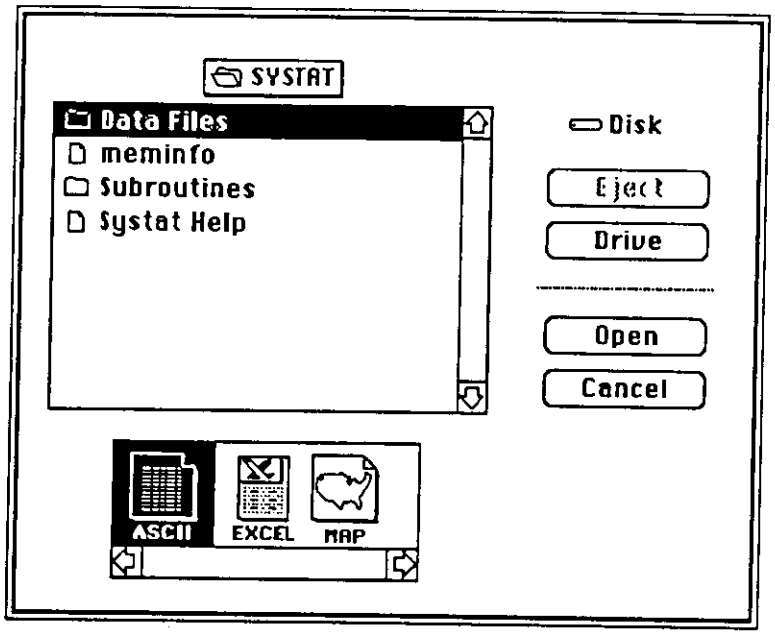
You can read text files saved from other applications with **Import...** from the **File** menu.

You must save the data from the other application as *plain text*. Most applications allow you to choose the file format in which to save with their **Save As...** dialog box. Choose [Text only] or a similar plain ASCII format.

- Open the other application
- Save the data file in a text only format
- Quit from the application

- Select **Import...** from the **File** menu
- Choose [ASCII]
- Select the file
- Click Open

© 1989, SYSTAT, Inc.



Note that after you click [Text], only text-format files are included in the selection list.

The file is translated and saved directly in a disk file. If you want to use the file right away, use **File/Open....**

Delimiters

The text files must separate values (columns) with tabs, commas, or spaces. These characters, which tell SYSTAT where one value ends and the next begins, are called *delimiters*.

Delimiters between values must be only a single character. Consecutive values *cannot* be separated by, say, both a comma and a tab. Use one or the other. See "Missing values" below.

Rows (cases) must be separated by carriage returns (paragraph marks). Most Macintosh applications can produce text files with these delimiters. The last row (case) of the data file must end with a Return; if it does not, SYSTAT will not read the last row of data.

Variable names

The file you open may include variable names on the first record (line) of the file.

© 1989, SYSTAT, Inc.

If the file does not have variable names, SYSTAT names the columns COL01, COL02, ..., COL*n*. You may name the variables before reading the file into SYSTAT, or change SYSTAT's generic variable names after reading the file.

Character values

Any character values that contain spaces or single or double quotation marks must be surrounded with (the opposite type of) quotation marks. Otherwise, character values do not need quotation marks, but you can use them if you want.

If your application does not produce text files that are precisely compatible with SYSTAT's requirements, you can use a word processor to fix them up. Remember to save in a plain text format from the word processor!

Missing values

If SYSTAT encounters two consecutive delimiters in a case, it assumes that a missing value appears in that position of the data file. SYSTAT then inserts a period (for numeric data) or a blank space (for character data).

You can also code missing values before importing the text file—again, use periods for missing numeric data and blank spaces surrounded by quotation marks for missing character data.

If missing values are coded some other way (for instance, an application might store missing values as –99), correct them before you attempt to do analyses on the data.

Reading Excel files

SYSTAT can automatically import data files from Microsoft Excel.

- Select **Import...** from the **File** menu
- Choose [Excel]
- Select the file
- Click Open

© 1989, SYSTAT, Inc.

Import... can read integer, number, character, label, and blank data types. Integer and number fields are converted to numeric variables for SYSTAT, blanks are read as missing values, and labels are converted to SYSTAT character variables. If any character values or variable names are longer than SYSTAT allows, they are truncated. If truncation results in duplication, an error results.

Import... cannot read encrypted or password-protected files. Unprotect them first.

Boolean and formula variables are not read.

Problems?

Some of my variables weren't read

Does the file contain variable names? If so, make sure that all the variables have labels and all the labels are in the first record (line, case, row) of the file.

My data got all messed up

Be sure that the text file doesn't have two delimiters adjacent to each other except to skip over missing values. Remember, only one delimiter can appear between each value. If SYSTAT encounters consecutive delimiters, it records a missing value; thus, the next value will end up one position "later" and so on. Use a text editor to fix the file before importing.

Reading map files

To create a map file, first create an ASCII file set up according to the description given in Appendix V of the *SYGRAPH* volume. Then, convert that ASCII file to a .MAP with **Import...** Note that the .MAP file must have a partner data file of the same name without the .MAP extension.

- Select **Import...** from the **File** menu
- Click [Map]
- Select your ASCII file of coordinates
- Click **Import**

© 1989, SYSTAT, Inc.

You can export data in two different ways:

1) You can **Cut** or **Copy** data into the Clipboard from SYSTAT, switch applications, and **Paste** the data into the other application.

2) You can save data from SYSTAT in a comma-delimited [Text] format or tab-delimited [Tabs] format with **Save as...**, and read the text file into the other application.

As for importing, it is quicker and easier to use the Clipboard to export relatively small amounts of data and to use external text files for larger datasets. Also, you need to use text files when exporting to a different platform, such as to a VAX.

Copying or Cutting into the Clipboard

You can transfer data from SYSTAT to another application through the Clipboard.

- Open the file in the Data Editor
- Select the range of data
- Select **Cut** or **Copy** from the **Edit** menu
- **Quit** from SYSTAT

- Open the other application
- Place the cursor where you want the data to appear
- Select **Paste** from the **Edit** menu

Saving text files

You can save SYSTAT data into plain text files with **Save as...**

- Select **Save as...** from the **File** menu
- Specify a filename and folder location
- Choose [Text] for comma delimiters or [Tabs] for tab delimiters
- Click Save
- Select **Quit** from the **File** menu

- Open the other application
- Read the text file according to that application's instructions

© 1989, SYSTAT, Inc.

Delimiters

Generally, [Tabs] is preferable for transferring data to another Macintosh application, unless that application instructs others.

Generally, [Text] (comma delimiters) is preferable for transferring data to an IBM-PC or mainframe application.

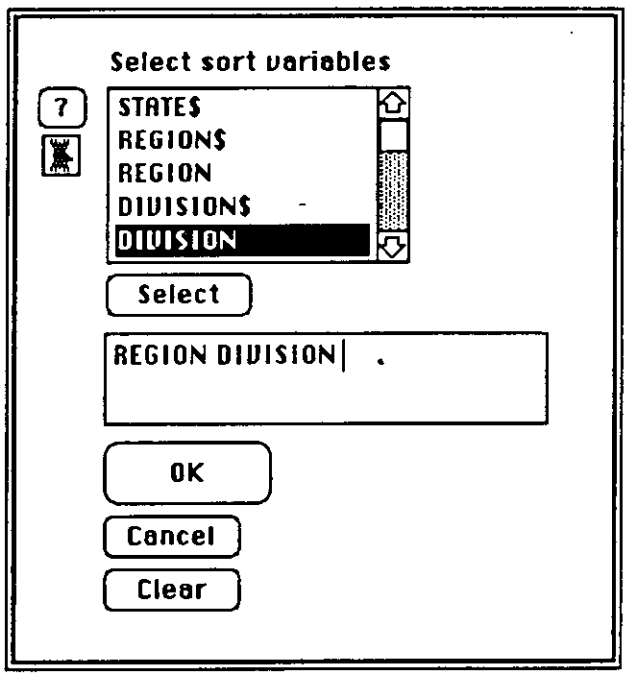
© 1989, SYSTAT, Inc.

SYSTAT lets you sort, rank, standardize, and transpose your data files with items in the **Data** menu. To use these items, you must have a data file open in SYSTAT Data Editor.

Sort...

Sort... orders all the cases in the file in ascending order according to up to ten numeric or character variables. The results are stored in a different file. You cannot change the current worksheet with **Sort...**

- Select **Sort...** from the **Data** menu
- Select the sorting variables, in order
- Click OK



After you click OK, a **Save as...** dialog box appears so you can specify a filename and folder location for the output sorted file.

If you do not select any variables, SYSTAT sorts using the first variable in the file.

© 1989, SYSTAT, Inc.

Nested sorting

If you select more than one sorting variable, SYSTAT does a nested sort. For example, if you select SEX$ and AGE, in that order, SYSTAT first sorts the males and females into two groups, and then sorts each group from youngest to oldest, so the SEX$ and AGE columns of your output file might look like this:

SEX$	AGE
female	5
female	8
female	9
female	10
male	6
male	6
male	7
male	9

Numeric order

SYSTAT sorts numeric data by putting missing values at the beginning of the file and the largest values at the end.

Character order

SYSTAT sorts character data in ascending ASCII order, with blanks (missing values) at the beginning. The sequence for an ascending character sort is:

```
 ! " # $ % & ' ( ) * + , - .
 / 0 1 2 3 4 5 6 7 8 9
 : ; < = > ? @
 A B C D E F G H I J K L M
 N O P Q R S T U V W X Y Z
 [ \ ] ^ _ `
 a b c d e f g h i j k l m
 n o p q r s t u v w x y z
 { | }
```

This means that words are sorted alphabetically with upper case words preceding lower case.

© 1989, SYSTAT, Inc.

Rank...

Rank... replaces values with their ranks in the variable you select. The results are stored in a different file. You cannot change the current worksheet with **Rank...**

- Select **Rank...** from the **Data** menu
- Select the rank variables
- Click OK

After you click OK, a **Save as...** dialog box appears so you can specify a filename and folder location for the output ranked file.

If you do not select any variables, SYSTAT ranks all the numeric variables in the file.

"Rank" is the case number a value would have if the dataset were sorted on that variable. For example, the smallest value would be rank 1, and the largest value in a fifteen-case variable would be 15. Ties are averaged. Here is a simple example that shows how ties are handled:

Value	Rank
1	1
2	2
3	3.5
3	3.5
10	5
11	6

Again, **Rank...** *replaces* the values of the variable with their ranks, so if you need to have both the ranks and the original values in one file, make a copy of the variable before ranking. **Rank...** changes the values only of the variable(s) you select, and all other variables are copied to the new file unchanged. You can use **Math...** to copy a variable, e.g. "Set AGE to AGE2."

Large files

Rank... does its work in memory. Therefore, you can run out of room with files containing thousands of cases. If this happens, first **Sort...** the file on the variable you want to rank. Then, use **Math...** to replace the values of the variable with CASE numbers. This method wont handle ties however.

© 1989, SYSTAT, Inc.

Standardize...

Standardize... replaces values of the variables you select with their sample standard scores, or *z*-scores. The results are stored in a different file. You cannot change the current worksheet with **Standardize...**

- Select **Standardize...** from the **Data** menu
- Select the variables to be standardized
- Click OK

After you click OK, a **Save as...** dialog box appears so you can specify a filename and folder location for the output standardized file.

If you do not select any variables, the first numeric variable in the file is standardized.

Standardize... subtracts the sample mean from each value and then divides the difference by the sample standard deviation. The standardized values have mean 0 and standard deviation 1.

Transpose...

Transpose... transposes the cases and variables of a file. That is, the variable columns become case rows and vice versa. The results are stored in a different file. You cannot change the current worksheet with **Transpose...**

Two restrictions apply. First, your file cannot contain character variables, except the first variable in the file may be a character variable containing names for variables in the transposed file. That variable must be called LABEL$. Second, you cannot transpose a file with more than 99 cases.

- Select **Transpose...** from the **Data** menu
- Specify an output filename and location
- Click Save

It is unnecessary to transpose a symmetric matrix (e.g., any matrix produced by **Corr**), because the transposed matrix is the same.

The transposed file contains one character variable called LABEL$ that contains the old variable names. The new columns (variables) are labeled COL01, COL02, ..., COL*n*, where *n* is the number of cases in the old file.

© 1989, SYSTAT, Inc.

6 Redirecting output

Saving output in files 112
Graphs 112
Save graph as... 112
Viewing PICT files 112
Text 113
Notes... 113
Save selected text as... 114
Results to file as... 114
Data produced by analyses 115
Using saved data 115
Printing output 116
Page preview... 116
Graphs 116
Print Graph... 116
Results to printer 117
Text 117
Page formats... 117
Print selected text... 118
Results to printer 118
Data produced by analyses 119
Placing output in the Clipboard 120
Graphs 120
Text 120
Data produced by analyses 120

© 1989, SYSTAT, Inc.

Redirecting output 6

Overview

SYSTAT lets you print and save analysis output, graphs, and data produced by analyses. You can print and save things one piece at a time, or you can tell SYSTAT to print or save everything. You can also place all three types of information in the Clipboard for transfer to other applications. (The previous chapter, **Data files**, discussed how to save, print, and place in the Clipboard SYSTAT data files.)

The easiest way to figure out what you're trying to print or save is by knowing what type of output appears in which window.

The View window is where SYGRAPH draws graphs. Some SYSTAT routines (in the **Stats** menu) also draw graphs in the View window—**Factor**'s factor loadings plots, for example. The View window shows *graphic output*.

The Analysis window is the window that shows text results—the results of operations selected from the **Stats** menu. When SYSTAT spits out a correlation matrix, for example, you are seeing *text output*.

SYSTAT and SYGRAPH have an optional command interface, which is discussed in Appendix I of the *SYGRAPH* and *SYSTAT* volumes. Any commands or messages that appear in the Command window are *text*.

Finally, you can save data produced by analyses. These are results that do not necessarily appear in any window, though some do appear in the Analysis window. You can save such things as means, minima, and maxima computed by SYSTAT. Or, you can save residuals from regression models in a data file. You can save a correlation matrix in a matrix-type data file. Generally, you save *data from analyses* by clicking the [Save file] option in the dialog box for that procedure. To learn more about saving data from a particular procedure, see the corresponding chapter in the *SYSTAT* manual.

© 1989, SYSTAT, Inc.

Graphs

Graphic output is anything that appears in the View window. You can save any graph in the View window into a PICT file. You can open PICT files with other Macintosh graphics applications such as SuperPaint. You can also bring graphs saved from a previous SYSTAT session back into the View window.

Note that you can transfer graphics to other applications two ways: either save a PICT file with SYSTAT and open that file with the other application, or copy a graph in to the Clipboard and paste it into the other application. See the "Placing output in the Clipboard" section for more details on the latter.

Note that the ACF, PACF, and CCF plots produced by **Series** items appear in the Analysis window and are considered to be text.

Save graph as...

You can save a graph *currently displayed in the View window* with **Save graph as...** from the **File** menu. If you have stacked several graphs in the View window, scroll to the graph you want to save and then select **Save graph as...** You can only save one graph at a time.

- Display a graph in the View window
- Select **Save graph as...** from the **File** menu
- Specify a filename and folder location
- Click Save

Viewing PICT files

Graphs are saved in PICT files. You can view a PICT file in the View window by double-clicking its icon, which launches SYSTAT and places the graph in the View window. You can then use the graph as usual—alter it with **Plot tools and colors...**, print it, etc.

You can also open PICT files with **Open...** from the **File** menu.

- Start SYSTAT
- Select **Open...** from the **File** menu
- Click [PICT]
- Select the PICT file you want
- Click Open

© 1989, SYSTAT, Inc.

To use PICT files with another Macintosh application, see the documentation provided with that application.

Text

Text is anything that appears in the Analysis window or the Command window.

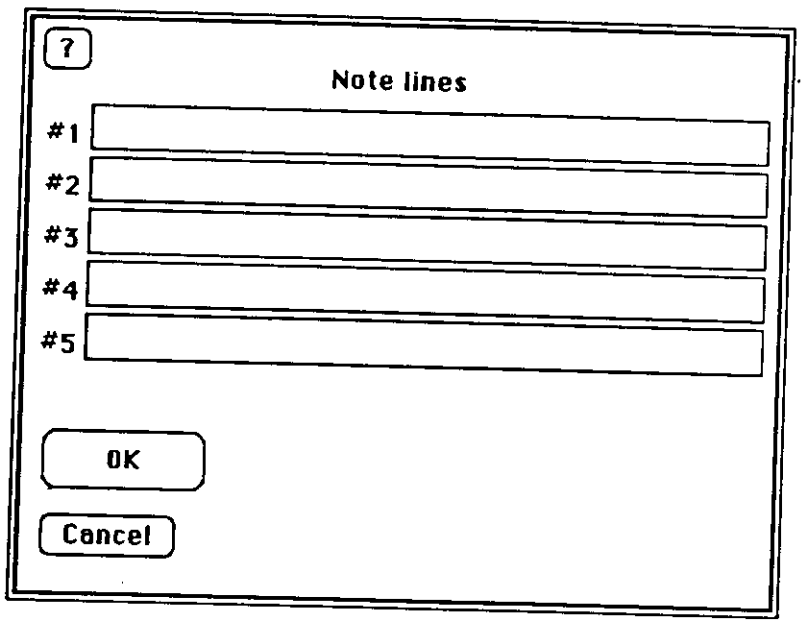
Both the Analysis window and the Command window are text editors. You can type in them and use **Cut**, **Copy**, and **Paste** in them. You can select all or part of a window's content for printing or saving. You can type a batch of commands in the Analysis window, **Copy** them into the Clipboard, and **Submit Clipboard** for processing.

Thus, you can use the Analysis window to compile quick reports: just type your introduction, comments and interpretations right into the output, select all the text, and save it or print it. You can change the font of the text with **Data/Formats...**

Notes...

You can also use **Notes...** to place up to five lines of comments in the Analysis window. The text you type in the lines with [#1]–[#5] is centered at the top of the Analysis window. Using **Notes...** is equivalent to using the mouse to position the cursor in the window and typing the notes.

- Select Notes... from the **Data** menu
- Type your notes in the boxes
- Click OK



© 1989, SYSTAT, Inc.

Save selected text as...

Save selected text as... saves the selected text in a file. First, highlight the text you want by using the mouse or **Select all** from the **Edit** menu. (Be sure that the window that shows the text is the active window by clicking on it.)

- Select the text you want to save
- Select **Save selected text as...** from the **File** menu
- Specify a filename and folder location
- Click Save

Note that the file type options are dimmed; this is because you can only save text in [Text] files.

If you name an already existing file, its contents are replaced with the new text. Be careful not to destroy previous work.

Results to file as...

Results to file as... saves *all subsequent* text output sent to the Analysis window. Output continues to appear in the Analysis window, and another copy goes to the file you specify.

Note that **Results to file as...** saves *subsequent* output: if you need to save some results already in the Analysis window, you'll have to produce them again after selecting **Results to file as...** (or use **Save selected text as...** to save them by hand to a different file).

- Select **Results to file as...** from the **Data** menu
- Specify a filename and folder location
- Click Save

To turn off **Results to file as...**, select **Results to window**. Otherwise, SYSTAT will save output until you quit.

© 1989, SYSTAT, Inc.

Data produced by analyses

You can save data produced by most **Stats** procedures in SYSTAT files by checking the [Save file] box in the dialog box. In some dialog boxes, the option has a different name, like [Save residuals].

- Select the procedure you want from the **Stats** menu
- Select the variables and options for the analysis
- Check the [Save file] (or similar option) box
- Click OK

Before SYSTAT starts its computations, it gives you the **Save as...** dialog box.

- Specify a filename and folder location
- Click Save

Then, you sometimes get another dialog box where you choose [Single] or [Double] precision and exactly what type of information to save. Make your choices and click OK.

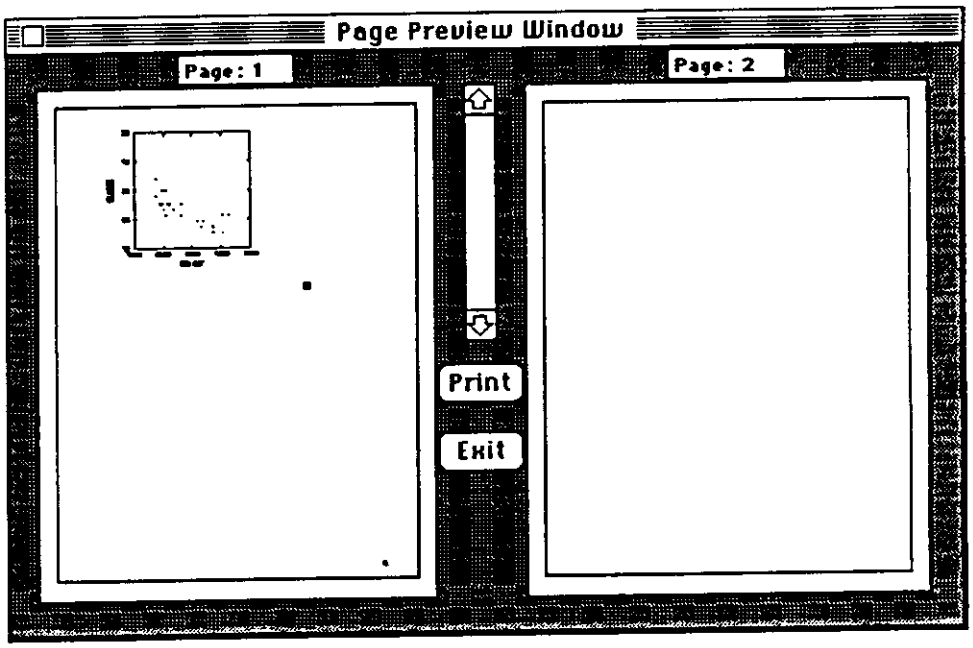
Using saved data

To examine the saved data, close the current data file (saving first if necessary) and open the file. You can click [Edit] if you want to see the data displayed in the Data Editor. Otherwise the file is just opened for further analysis or graphing.

© 1989, SYSTAT, Inc.

Page preview...

Page preview... from the **File** menu lets you see what you're about to print. After you select text or display a graph, select **Page preview...** from the **File** menu. You can move text and graphs, and you can also resize graphs using the gray resize box in the lower right corner of the graph marquee. The moving and resizing techniques are the same as for the ruler option, discussed in the **Graphing** chapter.



When you are ready to print, click [Print]. To cancel, click [Exit].

Graphs

You can print graphs two ways: one at a time with **Print Graph...** from the **File** menu, or all subsequent graphs ***and*** text with **Results to printer** from the **Data** menu.

Print Graph...

Print Graph... sends the graph currently displayed in the View window to the printing device.

© 1989, SYSTAT, Inc.

As with **Save Graph as...**, the View window must be active. Only the current graph is printed. If you want to print all the images stacked in the View window, you must print each one separately.

- Display a graph in the View window
- Make sure the View window is active
- Select **Print Graph...** from the **File** menu

Results to printer

Results to printer sends all *subsequent* graph *and* text output to the printing device. All output is still sent to the Analysis and View windows.

- Select **Results to printer** from the **Data** menu

Now do your analyses and graphs as usual. To turn printing off, select **Results to window** from the **Data** menu.

Text

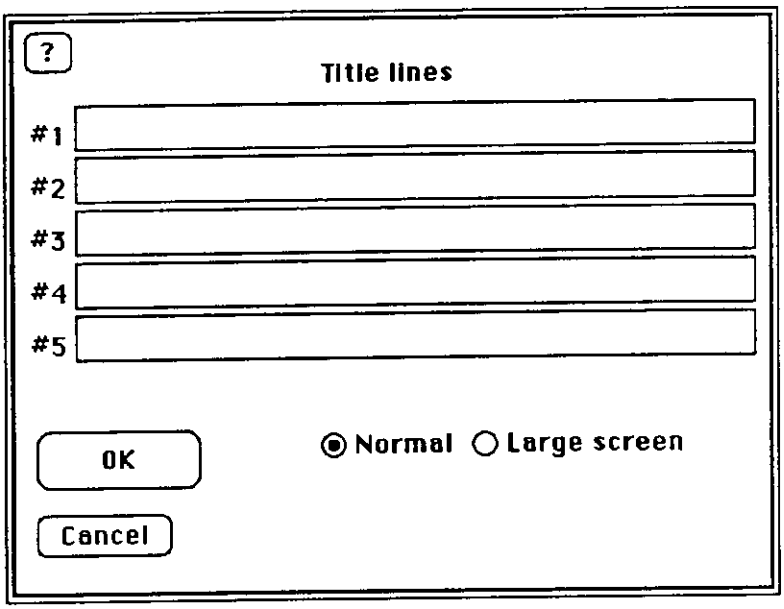
You can print text two ways: one piece at a time with **Print selected text...** from the **File** menu, or all subsequent output (text and graphs) with **Results to printer** from the **Data** menu.

Page formats...

Page formats... lets you specify how output is printed. First, you can enter up to five lines of title text to be displayed at the top of each page for printed text output.

You can also choose from two types of output: [Normal] or [Large screen]. These options control how long each line of text is. For instance, correlation matrices are usually too big to display in one piece, so some columns "wrap" to a subsequent set of lines. If you have a large screen (e.g., a standard Mac II screen), you can click [Large screen] so that more columns of text can be shown per line. Your choice of [Normal] or [Large screen] also affects the way output is printed and saved.

© 1989, SYSTAT, Inc.



If you want more than five lines of title text, use the command interface, discussed in **Appendix I** of the *SYSTAT* and *SYGRAPH* manuals. The PAGE command allows up to ten lines for its TITLE option.

Print selected text...

Print selected text... prints one piece of selected text. First, highlight the text you want by using the mouse or **Select all** from the **Edit** menu. (Be sure that the window that shows the text is active.)

- Select what you want to print
- Select **Print selected text...** from the **File** menu

Results to printer

Results to printer sends all *subsequent* text and graphic output to the printing device as well as the Analysis and View windows.

- Select **Results to printer** from the **Data** menu

Now do your analyses and graphs as usual. To turn printing off, select **Results to window** from the **Data** menu.

© 1989, SYSTAT, Inc.

Data produced by analyses

To print data saved from analysis procedures, first open the file in the Editor. Then, select **Print data...** from the **Editor** menu.

- Select **Open...** from the **File** menu
- Click [Edit]
- Select the file
- Click Open
- Select **Print data...** from the **Editor** menu

© 1989, SYSTAT, Inc.

Graphs

You can use **Copy Graph** from the **Edit** menu to copy the current graph into the Clipboard. You may then paste it into other applications.

Note that the View window must be active, or else **Copy Graph** does not appear in the **Edit** menu. **Copy Graph** only copies the graph currently displayed. You must copy and transfer other graphs stacked in the View window one at a time.

Text

To move text from the Analysis or Command windows into the Clipboard, use **Copy** or **Cut** from the **Edit** menu. First highlight the text you want by using the mouse or **Select all** from the **Edit** menu. (Be sure that the window is active.)

Data produced by analyses

To move data saved from **Stats** procedures into the Clipboard, first display the file in the SYSTAT Data Editor.

- Select **Open...** from the **File** menu
- Click [Edit]
- Select the file
- Click Open

Next, highlight the data you want (all or any rectangular range of cells in the Editor) by using the mouse or **Select all** from the **Edit** menu. Finally, use **Cut** or **Copy** to place the data in the Clipboard.

© 1989, SYSTAT, Inc.

© 1989, SYSTAT, Inc.

7 Analysis

Setting up 124
Formats... 124
Show options in effect 126
Record preferences 127

Using dialog boxes 128
Selecting variables and functions 128
Functions 128
Keyboard shortcuts 128
Options 129
Cancel 129
Clear 129
Saving dialog box settings 130
Variable bundles 130
Setting up bundles 131
Switching bundles 132

Analyzing subgroups 134
By groups... 134
Select... 135
Other ways to select subgroups 136

Weighting cases 137
Weight... 137
Cross validation 138
Saving analyses after weighting 138
Non-integer weighting 138

© 1989. SYSTAT, Inc.

Analysis 7

Overview

The first section of this chapter, "Setting up," discusses some ways you can adjust SYSTAT according to your preferences.

The second section, "Using dialog boxes," describes how the dialog boxes work and ways you can use them more efficiently. You can, among other things, save all of your dialog box settings from a particular session and restore them the next time you need to work on those data—so you hit the ground running in your next session. You can also set up "bundles" of variables so you only have to look at the variables that interest you, yet you can access the other ones whenever you need them.

The third section, "Analyzing subgroups" shows how to analyze data by groups through the use of grouping variables and how to analyze only variables that meet a given test condition (for instance, all the New England states but no states from other regions).

The final section, "Weighting datasets," shows how to weight cases in a dataset by a weight variable.

© 1989, SYSTAT, Inc.

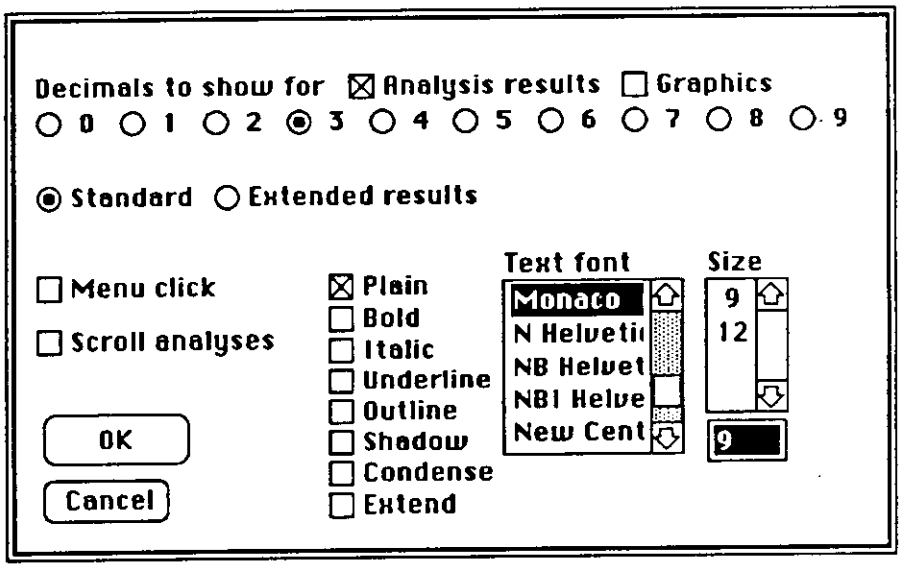
Here we discuss some options that can help make SYSTAT and SYGRAPH more efficient.

Formats... lets you control how SYSTAT displays text output. **Show options in effect** shows all the settings you've chosen for **Formats...** and **Page formats... Record preferences** stores your settings so that they will automatically be chosen whenever you start SYSTAT.

Formats...

Formats... controls the way SYSTAT displays text output.

- Select **Formats...** from the **Data** menu
- Make your choices
- Click OK



All **Formats...** settings stay in effect for the duration of your SYSTAT session unless you change them again. If you use **Record preferences** the settings are also chosen at the beginning of every subsequent SYSTAT session.

© 1989, SYSTAT, Inc.

Warning: in the *SYSTAT* volume, we occasionally ask you to change a **Formats...** setting for an example. It is understood that unless we state otherwise, all the *default* **Formats...** are used. We assume that if you try an example, you will restore the defaults as necessary. We usually do not remind you to change **Formats...** back. *If your output looks different from that shown in the manual, check your* ***Formats...*** *settings.*

[Decimals to show]

You can control the number of decimal places that are shown in analysis output. The default setting, 3, means that every number shown in analysis results is printed with three digits past the decimal. You can show from zero to nine places, and you can control [Analysis results] (text output) and [Graphics] (View window plots) separately:

- Select **Formats...** from the **Data** menu
- Choose the number of decimal places for [Analysis results]
- Click OK

- Select **Formats...** again
- Check the [Graphics] box
- Choose the number of decimal places for graphics
- Click OK

[Standard] or [Extended results]

For some SYSTAT procedures, you have a choice of [Standard] or [Extended results]. [Standard] results are the default and contain only the most common statistics. [Extended results] show some additional output for some procedures.

Your choice stays in effect until you change it or until you quit SYSTAT. If you only want extended results for one regression model, for example, switch to extended results before you estimate the model, estimate the model, and then switch back to standard results.

[Menu click]

If you check the [Menu click] box, you hear sound effects when you use SYSTAT's menus. These sounds might make it easier to know when you are moving through menus. If they merely annoy you, leave the option unchecked.

© 1989, SYSTAT, Inc.

[Scroll analyses]
Ordinarily, SYSTAT clears the Analysis window before it displays more results. If you prefer that all results accumulate, click [Scroll analyses]. New results are appended to the old, and the window scrolls.

One possible problem with scrolling analyses is that you may run into memory problems. If you do, just delete any text that you don't need any more. Close the Analysis window to clear it completely.

[Text font] and [Size]
SYSTAT displays text results in boldface 9 point Monaco by default. You may choose any font and size available on your machine; the selection will vary according to which fonts you have installed in your System file or in Suitcases. (See your *Macintosh System Software User's Guide.*) You may also choose from the various sizes.

Warning: you should generally choose a monospace (non-proportional width) font such as Monaco or Courier, because proportional width fonts (such as Helvetica and Times), where some characters are wider than others, destroy the neat columns of SYSTAT output. If you prefer to use a proportional font for a report, you will need to replace blank spaces with tabs with your text processor.

Show options in effect

Show options in effect from the **Goodies** menu tells what file is in use and what settings have been chosen.

```
SYSTAT options now in effect

   File in use is
   Disk:SYSTAT:Data Files:USDATA

   SYSTAT output display options are
      SYSTAT prints standard results
      in narrow page format, scrolls screen output, and displays

      Numeric output values print to 3 decimal places
```


© 1989, SYSTAT, Inc.

Record preferences

Record preferences from the **Goodies** menu records all the settings you have chosen in **Formats...** and **Page formats...** from the **Data** menu and **Formats...** from the **Editor** menu. It also records any changes you make to the size or locations of windows. Also, any windows that are open when you use **Record preferences** will be opened every time you start SYSTAT—even if the window is ordinarily not shown.

Your preferences are stored in a file called SYSTAT Pref in your SYSTAT folder. If you want to restore the defaults, simply drag SYSTAT Pref out of your SYSTAT folder.

© 1989, SYSTAT, Inc.

SYSTAT dialog boxes work pretty much the way you expect. However, they have some convenient new features.

Selecting variables and functions

To select a variable or function from a selection list, you may either double-click the variable or function, or click it (to highlight it) and then click the [Select] button beneath the selection list. If you use the latter method, be sure to use the [Select] button underneath the correct selection list.

Some dialog boxes have two variable selection lists. Usually the first (leftmost) list is the main selection list, and the second list is for selecting grouping variables.

Functions

When you select functions or operators, be sure that the argument of the function goes inside the parentheses. For instance, if you want ABS(RESPONSE), select ABS. Notice that the cursor is sitting inside the parentheses. If you now select RESPONSE, it will be placed correctly inside the ABS parentheses. You must be careful, however, when constructing more complex expressions. If you must, move the cursor with the mouse and type all or part of the expression from the keyboard.

Keyboard shortcuts

If you don't like using a mouse, you can use a number of keyboard shortcuts.

You can always type in the boxes where your selections are placed. You can edit selections as needed, type names of variables that aren't in the current bundle (see below), and type any symbols (such as + or -) that aren't available from selection lists.

Return for OK

You can press Return to set a procedure in motion; this is an alternative to clicking OK.

Moving from box to box

Some dialog boxes, like the **Recode...** box, have several boxes where expressions are entered. You can move between boxes with the mouse or by pressing the Tab key.

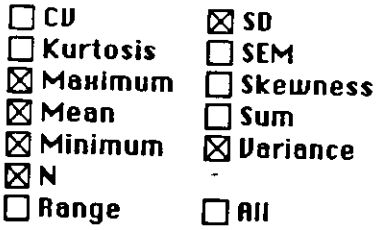
© 1989, SYSTAT, Inc.

Options

Most of the dialog boxes contain options. There are three types of options.

An *exclusive option* is one that prevents you from picking others; if you choose another, the first one turns off. Exclusive options have *radio buttons*. For example, many dialog boxes ask you to choose [Listwise] or [Pairwise] deletion of missing values. The radio buttons tell you that you can only pick one or the other. SYSTAT will not let you pick both.

A *nonexclusive* option is one that can be chosen in combination with others. Nonexclusive options have *check boxes*. You can click any of the options to put an X in the box (to "check" the box), and you can click a "checked" box to "uncheck" it or turn it off.



Nonexclusive options sometimes appear alone (like [Save file]), and you just decide whether they should be on or off.

A third type of option takes an *argument*. For instance, if you want confidence intervals for **Tables/Tabulate...**, you must specify the confidence level. To turn on the [Confi] option, type a value in the box.

Cancel

If you ever decide that you don't want to perform an analysis after all, just click Cancel instead of OK. SYSTAT closes the dialog box and lets you select some other procedure.

Clear

If you make mistakes selecting variables, you can either edit the box by hand or click Clear to erase all your selections and start over again.

© 1989, SYSTAT, Inc.

After you click OK to execute an analysis, SYSTAT "remembers" the choices you made in the dialog box: the variables you selected, the options you checked, etc. The next time you go to that dialog box, you see those same selections. You can click Clear to clear out the selection. To undo options, you must click them off.

If you use a dialog box to do one problem and then return to it to do a completely different problem, you will need to click [Clear] to remove earlier selections and look over the options to be sure you have what you want.

Saving dialog box settings

If you frequently do similar analyses on the same dataset in separate SYSTAT sessions, you may save time by restoring a previous day's dialog box settings. All the variable-selecting and option-clicking that you did before is already there, and you can save time if you need to do almost the same problems—you can just uncheck an option or select another variable and go.

- Select **Save as...** from the **File** menu
- Click [DLOG]
- Specify a filename and folder location
- Click Save

The next time you want to use those dialog settings, open that file:

- Start up SYSTAT
- Select **Open...** from the **File** menu
- Click [DLOG]
- Select the file you saved
- Click Open

You can save as many files of settings as you want, but you can only open one at a time. If you open a second DLOG file, the first file's settings are replaced in all dialogs by the second file,s settings.

Variable bundles

If you have a dataset with many variables, it may not be convenient to scroll through all the variable names before you come to the ones you want to select.

© 1989, SYSTAT, Inc.

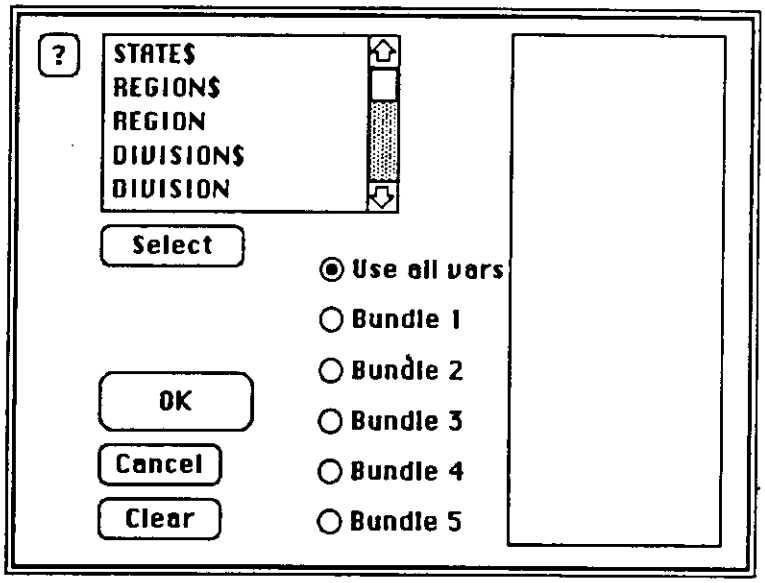
You can set up *variable bundles* to make variable selection easier. What you do is define a list of variables to belong to a bundle and then choose that bundle. Now, anywhere you would see a variable selection list, you will only see the variables in that bundle listed.

Setting up bundles

You can set up as many as five bundles at a time.

- Select **Define bundles...** from the **Goodies** menu
- Choose a bundle: click [Bundle 1], [Bundle 2], or whichever bundle you want to define
- Select the variables that you want to belong to that bundle
- Choose another bundle and select its variables, etc.
- Click OK

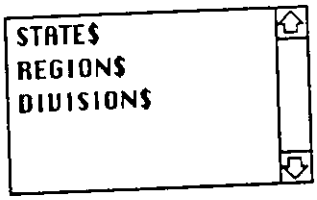
Notice that the tall box at the right lists all the variables selected for the current bundle.



The current bundle at the time you click OK becomes the current bundle for dialog boxes.

© 1989, SYSTAT, Inc.

Suppose you selected the three character variables, STATE$, REGION$, and DIVISION$, for Bundle 1. Now, when you select **Stats/Tables...**, for instance, you see those variables in the selection list:



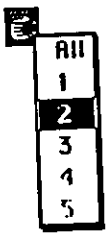
If you don't select any variables for a procedure, the usual defaults are either to use the first numeric variable in the file or all numeric variables in the file. If, however, you are a using a bundle, the defaults are either the first numeric variable in the *bundle* or all numeric variables in the bundle. In this respect, using variable bundles is a way of selecting subsets of *variables* rather than cases. (**Select...** selects subgroups of *cases* and is discussed in the next section, "Selecting subgroups.")

If you ever need to use a variable that does not belong to the current bundle, you can either type its name by hand in the selection box or switch bundles.

Switching bundles

You can switch bundles two ways. Either go back to **Goodies/Define bundles...**, choose a bundle, and click OK, or use the bundles icon in any dialog box.

The bundles icon in the upper left corner of any dialog box lets you change bundles. Press the mouse button and hold it on the icon. A list pops up. Select the bundle you want, or select [All] to use all the variables in the file.



© 1989, SYSTAT, Inc.

Keyboard shortcut

Another way to switch bundles from inside a dialog box is to hold the Command key (⌘) and type a bundle number. Type Command-0 (zero) to use all variables.

Saving bundles

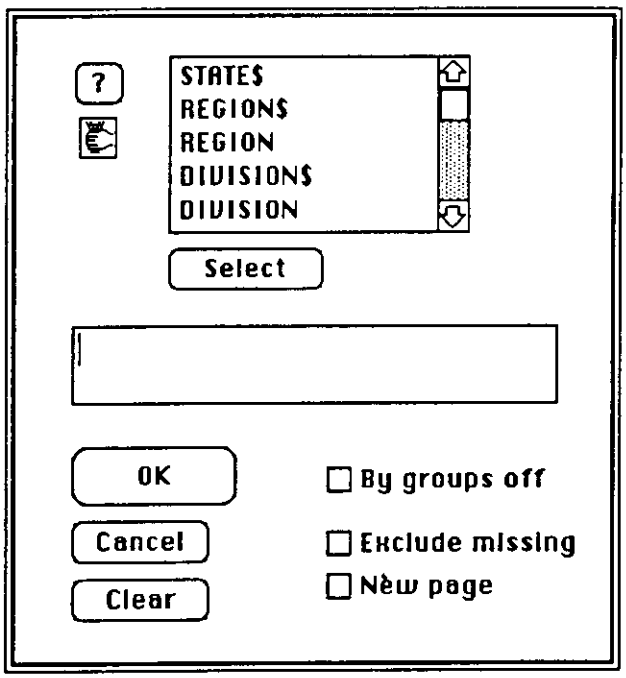
If you save a DLOG file (as discussed above), your bundles will be saved.

© 1989, SYSTAT, Inc.

By groups...

You can analyze subgroups with most SYSTAT procedures with **Data/By groups...**

- Select **By groups...** from the **Data** menu
- Select the grouping variable(s)
- Click OK



You may specify up to ten grouping variables. Your file must be sorted hierarchically on those variables. SYSTAT creates a new subgroup every time it encounters a new value on any grouping variable. Procedures perform analyses separately for each group.

See the **Data Editor** chapter under the "Special data structures" for information about grouping variables.

[By groups off]
Your grouping variables stay in effect until you quit from SYSTAT or open a new data file. You can turn them off by selecting **By groups...** again and clicking [By groups off].

© 1989, SYSTAT, Inc.

[New page]
The [New page] option causes SYSTAT to begin output for each by group on a new page for printed output.

[Exclude missing]
You can check the [Exclude missing] box to keep SYSTAT from analyzing cases that have missing values for grouping variables.

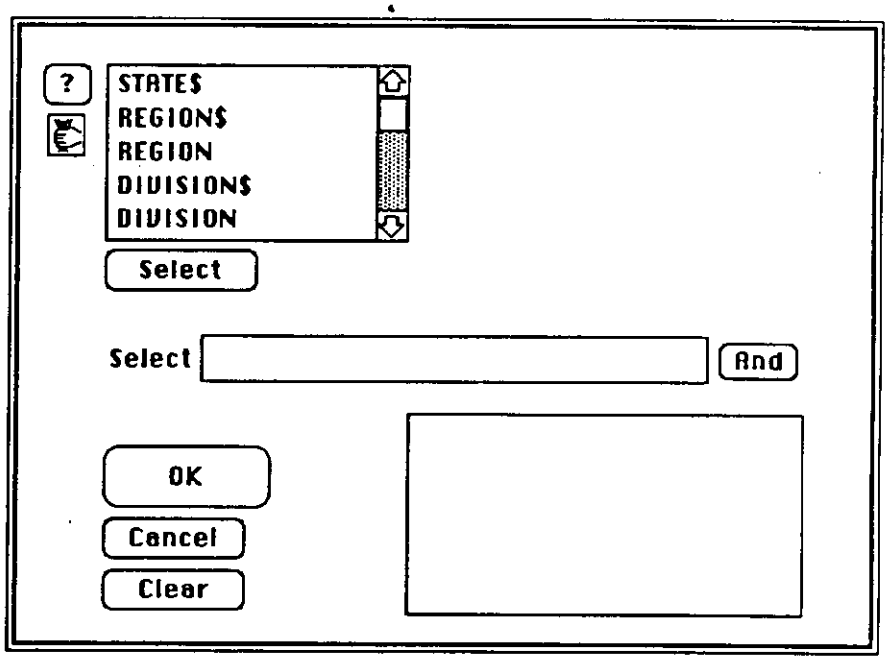
Select...

You can select a subgroup of cases from a file with **Select...**

- Select **Select...** from the **Data** menu
- Build a test expression by selecting variables and typing as necessary; use the [And] button for logical AND statements
- Click OK

Specify each condition as a numeric variable name followed by a relation followed by a number. The six possible logical relations are =, <>, <, >, <=, and >=. You may also test for character variables having a certain character value. Character values must be surrounded by single or double quotation marks. Character comparisons are case sensitive.

Use whatever combination of selecting, typing, and mousing necessary to build the test condition. If you need to use and, build one condition, click [And], build the next, etc. Each completed condition is moved to the large box in the lower right corner.



© 1989, SYSTAT, Inc.

You may specify up to ten conditions in a **Select...** statement. SYSTAT analyzes only those cases which meet ***all*** the conditions (i.e. you can use the logical [And], but [Or] is not available).

Turning selections off

Your selection stays in effect until you quit SYSTAT or open a new data file. To turn off the subgroup selection, select **Select...** again, click [Clear] to clear the test conditions from the box(es), and click OK.

With procedures that can save SYSTAT files (such as **MGLH**), you can use **Select...** to compute a model for some cases and then apply the model to other cases. For example, you can compute a regression on one set of cases and then use the estimated model to predict values on other cases. Just **Select...** a value on some variable for those cases to include in the model.

Other ways to select subgroups

The **Select...** item in the **Editor** menu is a more powerful version of **Data/Select...** First, **Editor/Select...** provides both [And] and [Or]. Second, if you select cases using **Editor/Select...**, you can later extract the cases to a separate data file using **Editor/Extract selected...**

The scatterplot brushing tools let you select cases visually from scatterplots. To use scatterplot brushing tools, you must have both the Editor and View windows open, and (obviously) you can only work from a scatterplot. Since the Editor is open, you can also use **Extract selected...**

Data/Select... is only preferable to the other methods when your data file is bulky, since the Data Editor can be slow for large files. Both other methods are in the **Editor** menu, which is only present when the Editor is open.

The three ways to select cases are compared in the **Data Editor** chapter. **Editor/Select...** is discussed in detail in that chapter also. The scatterplot brushing tools are introduced in the **Exploratory data analysis** chapter, later in this volume.

© 1989, SYSTAT, Inc.

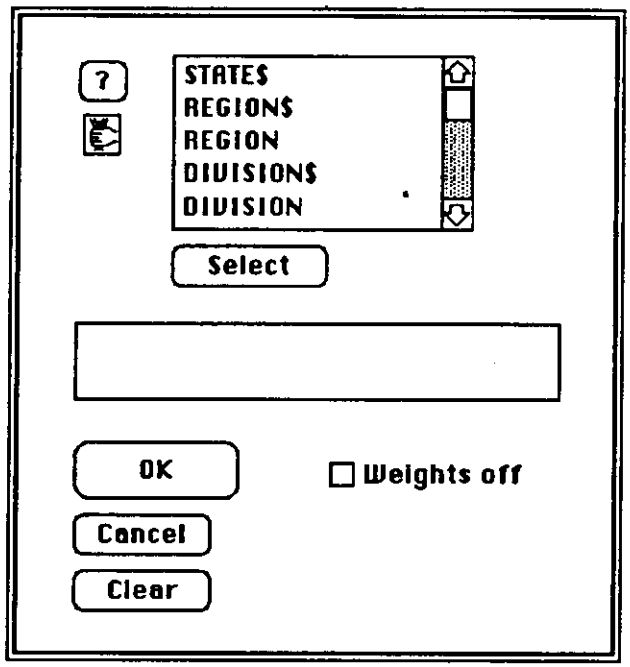
Weighting replicates the cases in a dataset according to a given weight variable. You can use case weighting to "unpack" frequency distribution datasets by weighting by a COUNT variable. (Frequency distributions are discussed in the **Data Editor** chapter in the "Special data structures" chapter.)

Weight...

Weight... allows you to specify a weighting variable for statistical analyses. SYSTAT replicates each case the number of times specified by the weighting variable.

- Select **Weight...** from the **Data** menu
- Select a weighting variable
- Click OK

Any analyses you do subsequently will use the weighted version of the dataset.



© 1989, SYSTAT, Inc.

For each case, SYSTAT truncates the value of the weighting variable to an integer. SYSTAT uses the sum of the weights as the sample size. If the value of the weighting variable is less than 1 for a given case, SYSTAT excludes that case from analysis.

Cross validation

In some procedure modules, such as MGLH, you can use weights of one to identify the cases you wish to use to estimate a model and weights of zero for cases for which you only want predictions or classifications.

Saving analyses after weighting

Statistical modules which can produce SYSTAT files save the unweighted number of cases after an analysis, i.e. the same number of cases in the input file. This way, you can merge results with the original data file without additional work.

[Weights off]

Weighting stays in effect until you quit SYSTAT or open a new data file. To turn weighting off, select **Weight...** again and click [Weights off] and OK.

Non-integer weighting

This method of weighting is different from some programs that allow weighting of variables but do not change the total sample size. Often, this type of weighting involves non-integer weights. To do this type of weighting in SYSTAT, use DATA to multiply the data by the weighting variable. This is the type of weighting done with the SAS WEIGHT option (as opposed to FREQ). The DATA facilities are discussed in detail in the *DATA* volume, but here is a brief example:

```
DATA

USE CAPITAL
SAVE WCAPITAL
LET INCOME=INCOME*ABILITY
RUN
```

The example weights INCOME by the product of INCOME and ABILITY and saves the results in the file WCAPITAL.

© 1989, SYSTAT, Inc.

© 1989, SYSTAT, Inc.

8 Graphing

Sizing and positioning graphics 142
The ruler tool 142
Positioning graphs by mouse 143
Resizing by mouse 144
Locating by typing 144
Resizing by typing 144
Example 145
Resize box 148
Moving graphs 148
Resizing graphs 148
Resizing example 148
Commands 149

Stacking and overlaying 150
Overlay, stack, and next 150
Overlay plot 150
Stack plot 150
Next plot 151
Movie... 151
Time series movie example 152

Options 157
Color... 157
Formats... 157
Global settings... 157

Drawing tools 158
Plot tools and colors... 158
Rectangle 159
Rounded-corner rectangle 159
Oval 159
Circle 160
Line 160
Text 160
Pointer 161
Colors 162
Hints 162

© 1989, SYSTAT, Inc.

Graphing 8

Overview

The first section of this chapter is about "Sizing and positioning graphics" in the View window. You can control the size and location of graphs in the View window by using the ruler option in every **Graph** menu dialog box. You can also move and resize graphs once they are drawn.

The next section discusses "Stacking and overlaying graphs" in the View window. Whenever you already have a graph in the View window and you ask SYGRAPH to plot another, you have a choice of laying the new one over the previous one, or scrolling the first one up in the window and drawing the new one on a new "page" in the window, or clearing out the View window and drawing the new one.

The "Options" section introduces some global settings for graphs that are discussed in detail in the **Common options** of the SYGRAPH manual. Here, we just introduce them so you know what the possibilities are; you can skip ahead to *SYGRAPH* manual whenever you're ready.

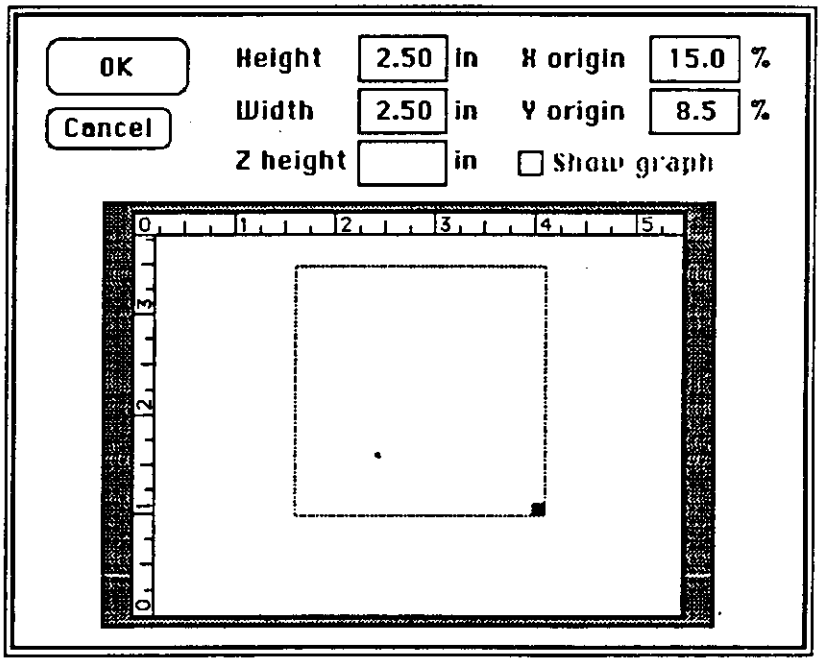
The "Drawing tools" section introduces SYGRAPH's new MacDraw-like tools for preparing presentation graphics. SYGRAPH's **Plot tools and colors...** are object-oriented text and drawing tools that let you add shapes, lines, and text in full color.

© 1989, SYSTAT, Inc.

The ruler tool

Every plot dialog box has a ruler icon alongside its palette of options. When you click this ruler, a page preview-style dialog appears. This dialog box lets you adjust the height, width, and position of the graph-to-be visually or by typing measurements.

- Click the ruler icon
- Specify size and position visually or by typing numbers
- Click OK



The dialog box shows a marquee representing the size and location in the View window of the graph that will be drawn. The marquee shows only the *graph* size—it does *not* contain any scales, labels, titles, and so on. The *graph* is strictly the axes and drawn portions of a graph.

The gray box in the lower right corner of the marquee is a *resize box*. You can drag the resize box to change the size and shape of the marquee (and thus the graph).

© 1989, SYSTAT, Inc.

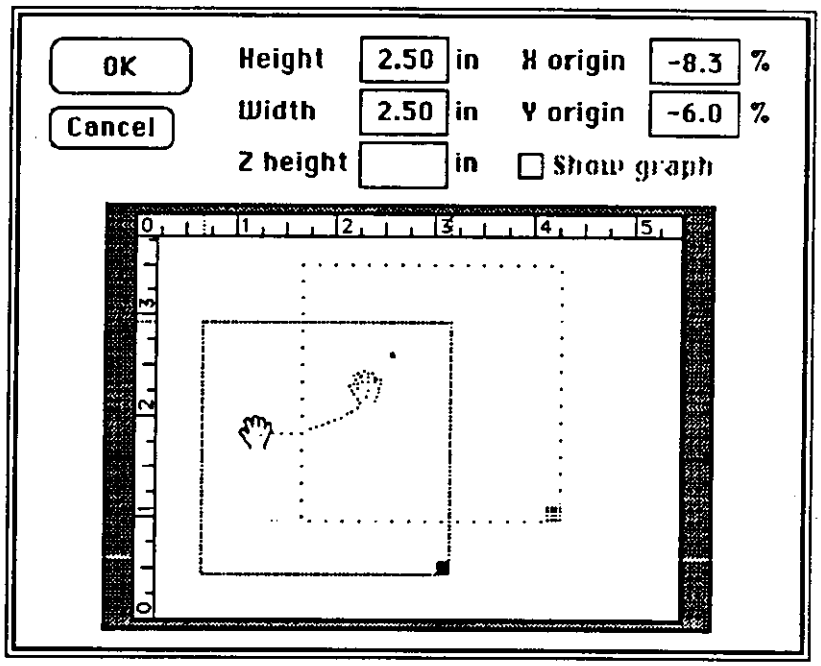
As you can see, SYGRAPH's default is to place a two and one-half inch square graph in upper center of the View window. The rulers along the top and side show the current size of the View window. If you change the size of the View window, the ruler dialog box changes accordingly.

[Show graph]

If you already have a plot on display in the View window, you can click [Show graph] to see a sketch of the plot in the screen. This enables you to position subsequent graphs relative to previous ones easily. With [Show graph] you can ensure that overlayed graphs and stacked graphs will align neatly.

Positioning graphs by mouse

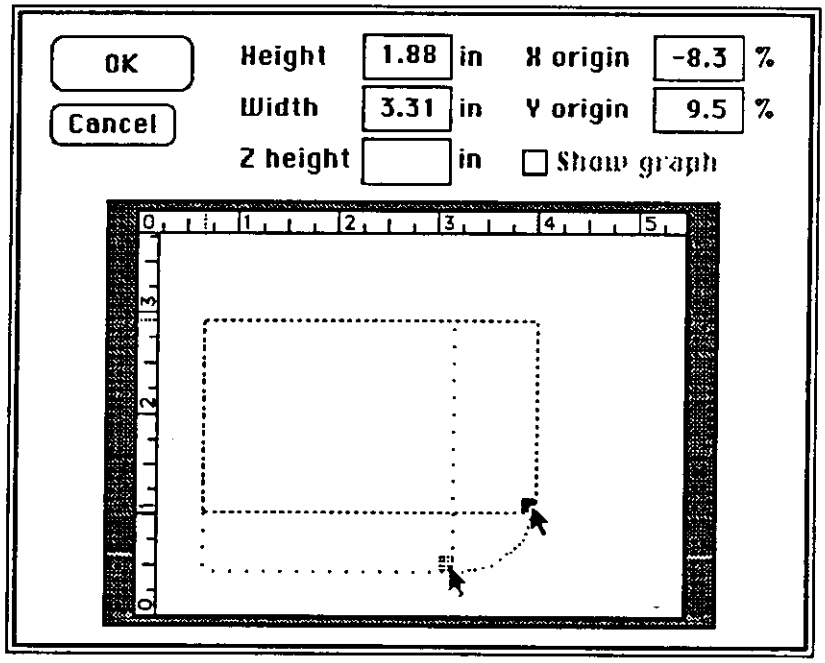
To change the location in the View window of a subsequent graph, position the pointer inside the marquee. Notice that the pointer is now represented by a hand. Press the mouse button, drag the marquee to its new position, and release the button.



© 1989, SYSTAT, Inc.

Resizing by mouse

To change the size and shape of the marquee with the mouse, position the pointer over the gray resize box in the lower right corner of the marquee. Press the mouse button, drag the resize box to another location, and release the button.



Positioning by typing

You can position a graph in the View window by typing coordinates for the [X origin] and [Y origin] options. The coordinates you specify determine the location of the lower *left* corner of the graph.

Notice that the coordinates change automatically when you use the hand to drag the marquee.

The units are percentage of the total current size of the View window, and the coordinates measure the distance up and over from the spot near the lower left corner of the View window that SYGRAPH considers to be (0,0). You can find where (0,0) is for a View window of a particular size either by moving the marquee around and watching the coordinates or by specifying 0 for [X origin] and [Y origin].

Resizing by typing

You can specify graph size in inches by typing values in the [Height] and [Width] boxes. You can also specify a [Z height] for 3-D plots.

© 1989, SYSTAT, Inc.

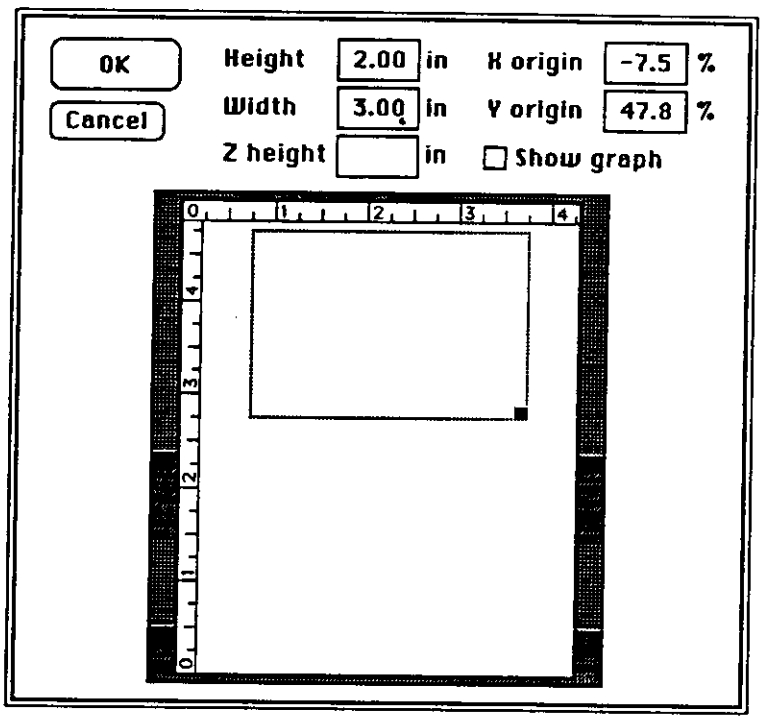
Example

We are going to use the ruler to place to density plots in the View window: a fuzzygram of average WINTER temperatures from the US dataset, and a frequency polygon of the same variable. We want to stack them vertically so we can compare the information that the two plots show.

First, we must resize the View window so that it will be large enough to accommodate two 2" by 3" plots. If the View window is not already open, select **Show view window** from the **Goodies** menu.

Now, use the resize box in the lower right corner of the window to make the window about 4 inches wide and 6 inches tall. (If you have a smaller screen, just adjust all these sizes as needed.) Now we're ready.

- Select **Density/Fuzzy** from the **Graph** menu
- Select a variable: WINTER
- Click the ruler icon
  - Use the resize box to make the marquee 2 inches tall and 3 inches wide (watch the values in the [Height] and [Width] boxes
  - Use the hand to move the marquee to the top, center part of the screen
  - Click OK
- Click OK

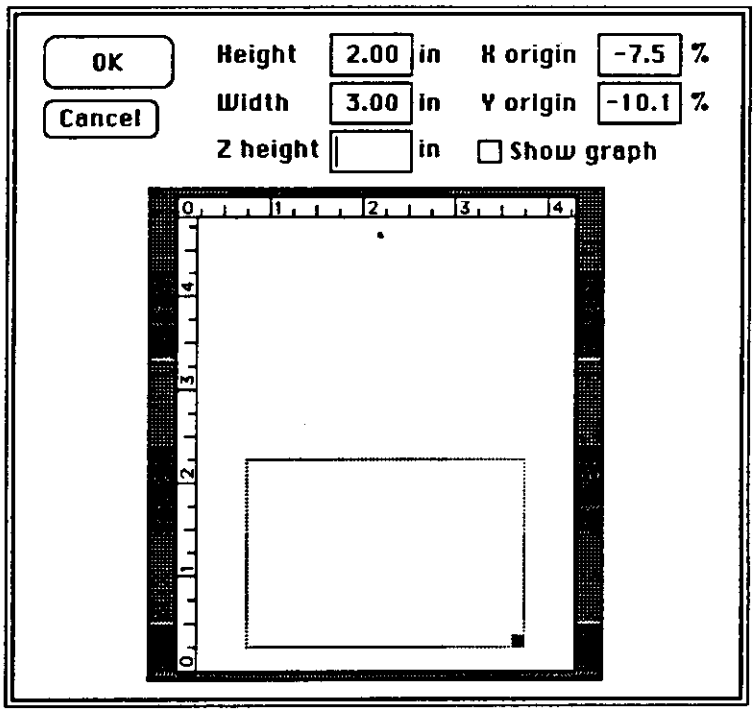


© 1989, SYSTAT, Inc.

SYGRAPH draws the fuzzygram. If your fuzzygram doesn't come out exactly where you wanted it, pick it up and move it in the view window. (Instructions for that appear in the next section, "The resize box.") When you're satisfied, plot a frequency polygon in the lower part of the window.

- Select **Density/Polygon** from the **Graph** menu
- Select a variable: WINTER
- Click the ruler icon
  - Use the resize box to make the marquee 2 inches tall and 3 inches wide (watch the values in the [Height] and [Width] boxes
  - Use the hand to move the marquee to the bottom, center part of the screen
  - Click OK
- Click OK

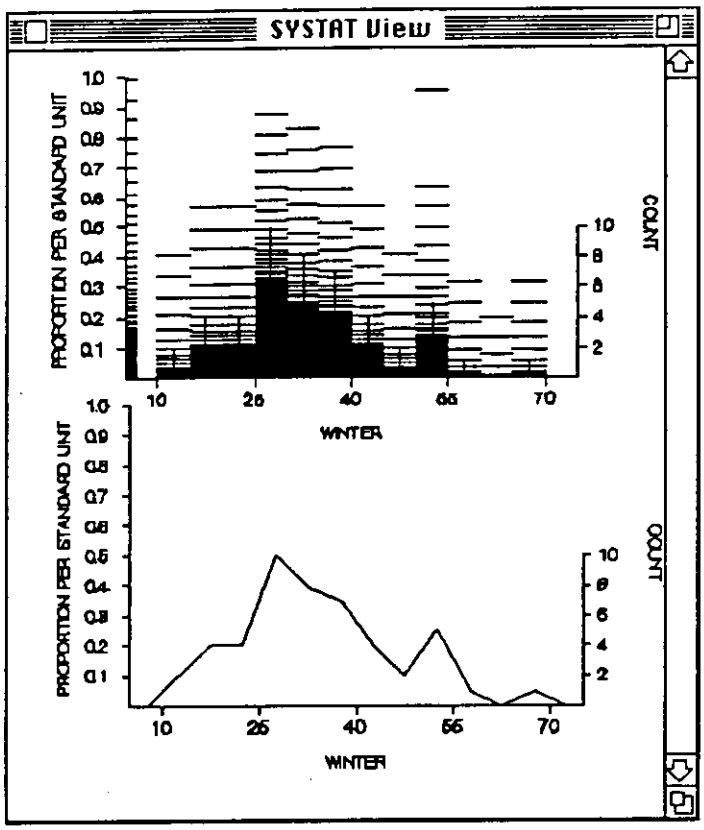
Note that in this step you can check the [Show graph] box to make sure you are positioning the next graph correctly. But don't worry—if you are a little off, you can move the graph in the View window to fix it.



© 1989, SYSTAT, Inc.

Before SYGRAPH draws the next plot, it displays an Overlay plot/Stack plot/Next plot. We discuss this dialog box in the second part of the chapter, "Stacking and overlaying." For now, just click [Overlay plot]; this means that you want the graph to appear on the same page as the fuzzygram.

If you are not satisfied with the position of the plot, just pick it up and move it. Here's how our View window looked when we had gotten everything adjusted to our satisfaction:



© 1989, SYSTAT, Inc.

Resize box

After a graph is plotted in the View window, you can move it and resize it visually. (If you have a scatterplot brushing tool selected, first click it off before attempting to move or resize the graph.)

Moving graphs

To move a graph, just position the pointer somewhere inside the graph. When you press the mouse button, a marquee appears around the graph. The graph (everything inside the marquee) is an *object*, and you can move the entire object. You can read more about objects in the "Drawing tools" part of this chapter.

Thus, to move the graph, just press the mouse button inside the graph, drag the graph to a new location, and release the mouse button.

Resizing graphs

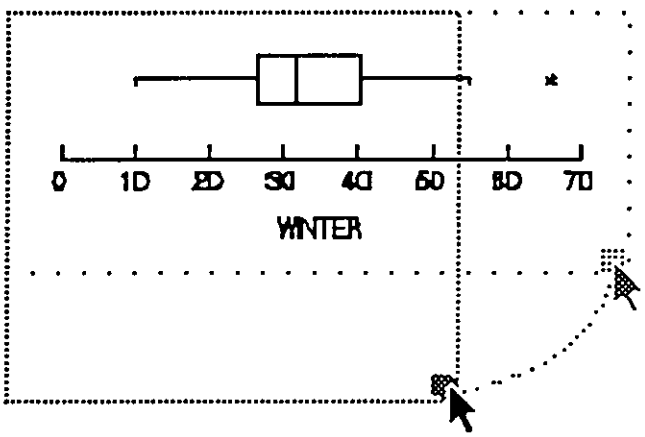
The light gray box in the lower right corner of the graph marquee is a resize box. It works the same way as the resize box you saw in the ruler dialog box. Just press the mouse button on the box, drag it to a new location, and release the button. The graph is resized accordingly.

Resizing example

Here we draw a box plot with SYGRAPH's default size and then resize it in the View window. We are using the US dataset.

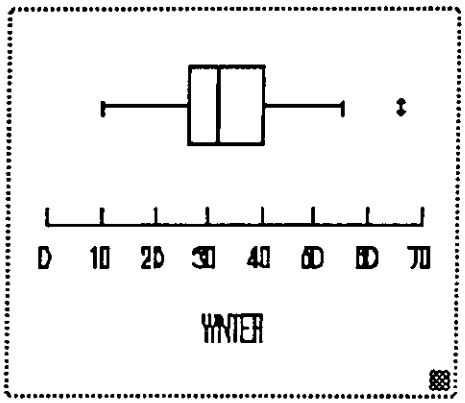
- Select **Box** from the **Graph** menu
- Select a variable: WINTER
- Click OK

After the plot is drawn, click on it to show the marquee and resize box. Then, drag the resize box to a new location.



© 1989, SYSTAT, Inc.

The result is a box plot that looks like a Cinemascope movie on TV.



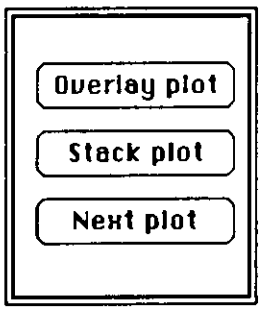
Commands

If you choose to use the optional command interface for SYGRAPH, you can use the HEIGHT and WIDTH (and ZHEIGHT) commands to adjust graph size. See **Appendix I: Command reference** in the *SYGRAPH* volume for information about SYGRAPH commands.

© 1989, SYSTAT, Inc.

Overlay, stack, and next

You might want to superimpose one graph onto another or place two graphs side by side to compare them. Whenever there is a graph in the View window and you ask for another graph, SYGRAPH displays the following dialog box:



Overlay plot

If you want to superimpose a new graph onto the one already in the View window, click [Overlay plot]. SYGRAPH draws the next plot directly on top of the current plot. If you want the next graph in some other location, use the ruler to trace the new size and position before clicking OK to start drawing the graph.

See the "Sizing and positioning graphics" section above for an example of overlaying.

Stack plot

[Stack plot] scrolls the current graph upward in the View window before drawing the next plot. The vertical scroll bar becomes active, and you can scroll up and down through all the plots that the window accumulates. Once you have stacked plots in the window, you can use **Goodies/Movie...** to view them in quick succession.

If you submit a batch file of commands (see the **Command reference** in the *SYGRAPH* manual) that draws several graphs, SYGRAPH automatically stacks them; you don't need to click the [Stack plot] button yourself. If you want to overlay graphs, place them between BEGIN and END statements.

© 1989, SYSTAT, Inc.

Next plot

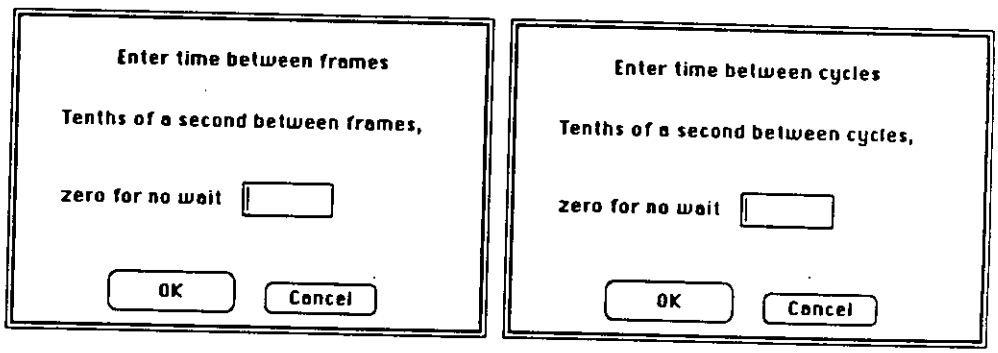
[Next plot] clears all previous graphs from the View window before drawing the next. If you have stacked plots, *all* the plots are removed. The vertical scroll bar is disabled since you have only a single plot in the View window.

Movie...

You can view a stack of graphs in quick succession for a movie-like effect. Movies are especially useful for monitoring the effects of a series of transformation, extractions, smoothings, or weighting. Astronomers use a "flicker box" to reveal objects that move slightly from time to time. You can do this with two scatterplots in a Movie.

- Display a plot
- Do another plot, and use [Stack plot] to scroll the previous plot upward in the View window
- Continue stacking
- Select **Movie...** from the **Goodies** menu
- Specify the wait period between graphs in tenths of a second and click OK
- Specify the wait period between cycles in tenths of a second and click OK

To stop the movie, press the mouse button and hold it briefly, until the movie stops.



You can resume stacking graphs after viewing a Movie, and you can do another Movie any time. You do not lose the graphs in the View window until you close the window or click [Next plot]. **Movie...** does not clear the window.

Low or high resolution?

You can do movies with high resolution (PostScript) graphs (use **Options/Global settings...**), but since high resolution graphs draw quite slowly, we do not recommend it.

© 1989, SYSTAT, Inc.

Time series movie example

In the **Series** chapter of the *SYSTAT* manual, we discuss the 4253H filter of Tukey.

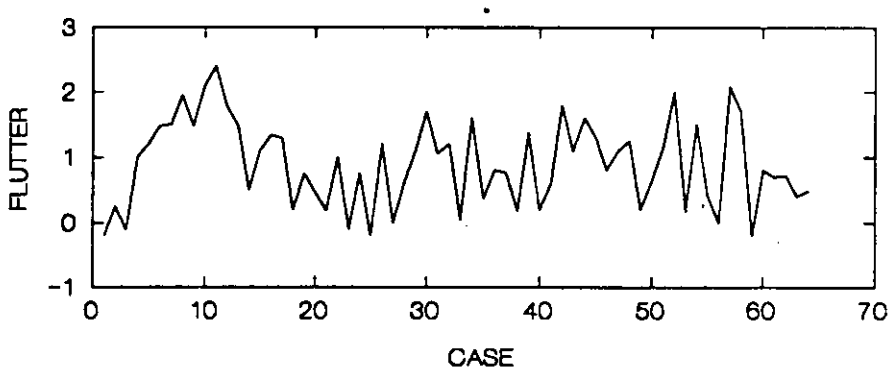
- Running median smoother, window 4
- Running median smoother, window 2
- Running median smoother, window 5
- Running median smoother, window 3
- Running means smoother, window 3, weights 1, 2, 1

A movie can show beautifully how the smoother gradually tames the series. Since we can't print a Movie in the manual, you'll have to try this yourself. We're using the AIRCRAFT dataset. After each smoothing, we'll do a time series plot.

First, here's a plot of the original series:

- Select **Series/Case plot...** from the **Stats** menu
- Select the variable: FLUTTER
- Set [Lag] to 64
- Click OK

(If you had something in your View window already, click [Next plot] to start a fresh stack.)



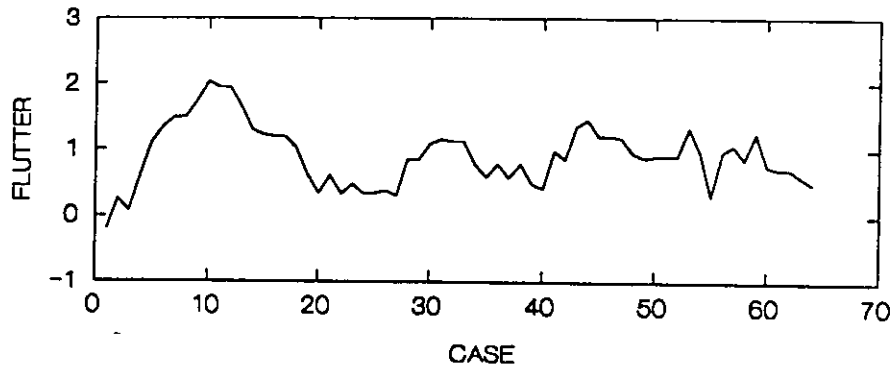
Now we smooth and plot again. Since the vertical scale values are likely to change after we tame the series, we set [Min] and [Max]. This way, all of the plots will be on the same scale and will be easily comparable.

© 1989, SYSTAT, Inc.

- Select **Series/Smooth...** from the **Stats** menu
- Select the variable: FLUTTER
- Choose [Median] smoothing
- Set [Median window=] to 4
- Click OK

- Select **Series/Case plot** from the **Stats** menu
- Set [Min] to –1 and [Max] to 3
- Click OK

We click [Stack plot] and the first smoothing is displayed.

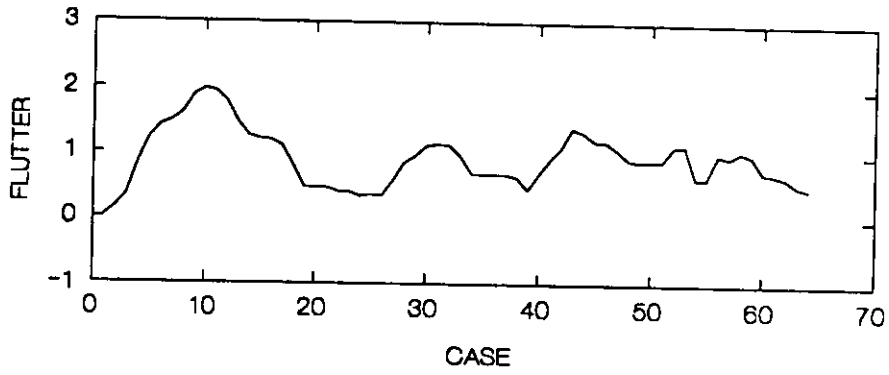


Now, we do the next smoothing and plot again. We need only change the smoothing window size, since the variable and type of smoothing are the same. We don't need to do anything in the **Case plot** dialog box since we want all the same settings.

- Select **Series/Smooth...** from the **Stats** menu
- Change the [Median window=] to 2
- Click OK

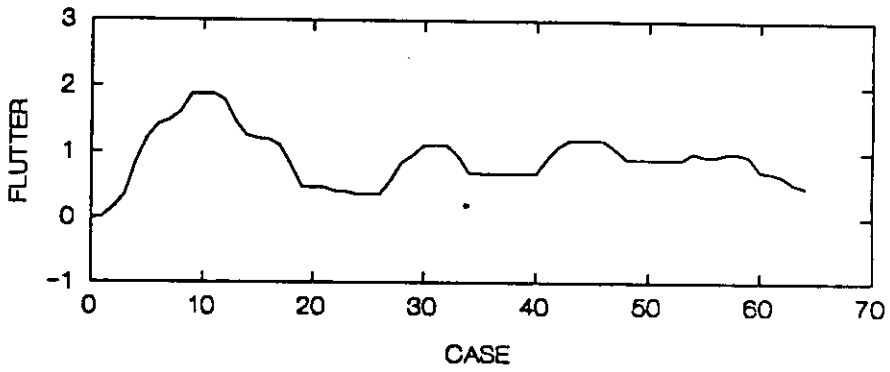
- Select **Series/Case plot** from the **Stats** menu
- Click OK

© 1989, SYSTAT, Inc.



- Select **Series/Smooth...** from the **Stats** menu
- Change the [Median window=] to 5
- Click OK

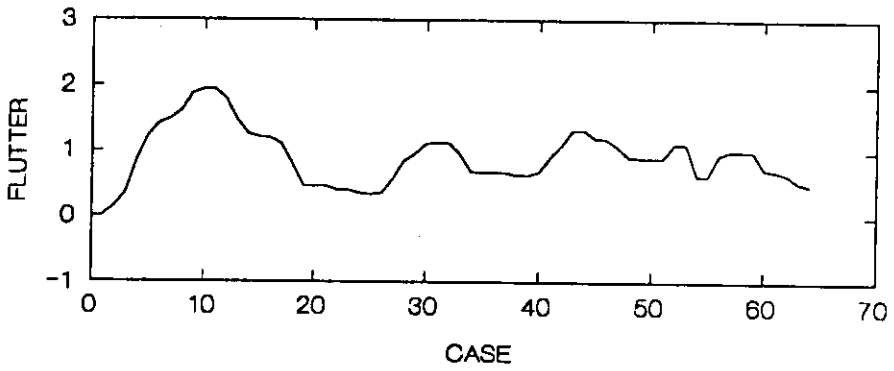
- Select **Series/Case plot** from the **Stats** menu
- Click OK



- Select **Series/Smooth...** from the **Stats** menu
- Change the [Median window=] to 3
- Click OK

- Select **Series/Case plot** from the **Stats** menu
- Click OK

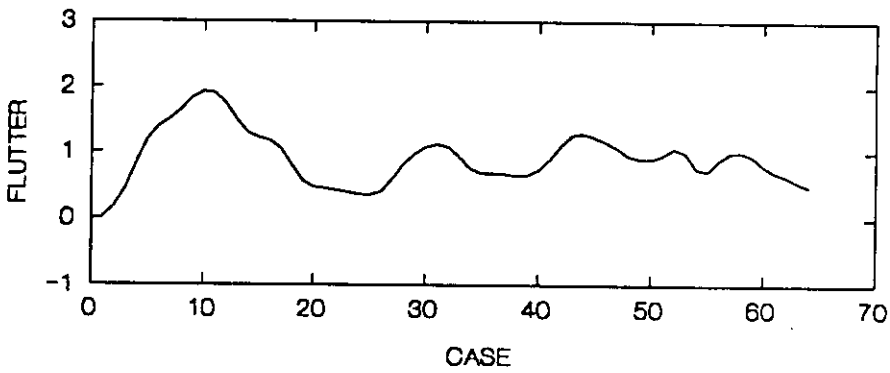
© 1989, SYSTAT, Inc.



Now, the final Hanning step.

- Select **Series/Smooth...** from the **Stats** menu
- Choose [Mean] smoothing, window 3
- Set [Weight] to 1, 2, 1
- Click OK

- Select **Series/Case plot** from the **Stats** menu
- Click OK



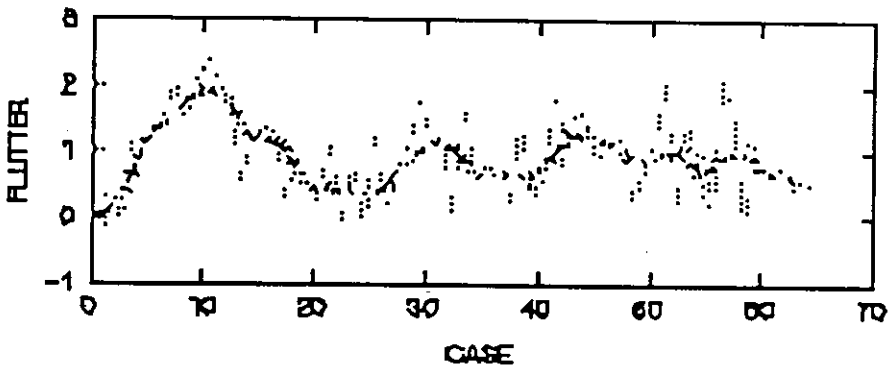
© 1989, SYSTAT, Inc.

We're all set: we have the original series and five smoothed versions all plotted and stacked in the View window. It's time for a movie.

- Select **Movie...** from the **Goodies** menu
- Specify 0 for time between graphs and click OK
- Specify 10 for time between cycles and click OK

SYGRAPH should now flip through your stack of graphs, pausing for a full second before starting back at the original series. When you have seen the Movie enough times, press and hold the mouse button until it stops.

A sextuple-exposure shot of the Movie is the best we can do here...



Popcorn is optional.

© 1989, SYSTAT, Inc.

Options

The **Options** item in the **Graph** menu has three items for controlling graph features globally—that is, for setting conditions that are in effect for all graphs, all the time, always. (Until you change them or quit.) All three items are discussed in greater detail of the **Common options** chapter of the *SYGRAPH* manual. We merely mention them here so you're aware of them.

Color...

With **Color...** you can choose the color settings to be used for all subsequent plots. You can choose colors for the "Background," "Foreground," "Graph," and "Labels." Definitions and instructions for these appear in the **Common options** chapter of the *SYGRAPH* manual.

Formats...

Formats... lets you control placement of plots being overlayed in three-dimensional perspective; scaling to expand or reduce plots; color separations; line thickness; and fonts, sizes, and styles. All these are discussed in the **Common options** chapter.

Global settings...

Global settings... offers three on-off switches that control the appearance of graphs.

[Automatic graph font scaling] automatically resizes text in graphs when you resize graphs: if you make the graph bigger, SYGRAPH makes the text bigger to match, and the reverse.

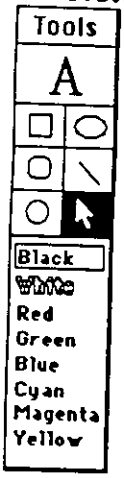
[Graph scales as editable text] lets you control whether the scales (the numbers marking axes) can be edited with the **Plot tools and colors...** or whether they are fixed to the graph and unchangeable.

Finally, [High resolution] controls whether graphs are drawn with Quickdraw commands (low resolution) or PostScript commands (high resolution).

© 1989, SYSTAT, Inc.

Drawing tools

Plot tools and colors...



You can draw and write on standard resolution graphs in the View window using SYSTAT's plot tools. The tools will be familiar to users of MacDraw and similar object-oriented graphics packages for the Macintosh.

- Display a graph or open a PICT file
- Select **Plot tools and colors...** from the **Goodies** menu
- Click the View window to make it active
- Choose a tool by clicking it

You may use both scatterplot brushing tools (discussed in the next chapter, **Exploratory data analysis**) and drawing tools on a given graph. You can draw on the graph with any of five tools: rectangle, rounded-corner rectangle, circle, oval, and line. You can add text with the text tool (A) and edit any of the objects (including the graph itself) with the pointer. You can change the color of any object.

Plot tools are not available with [High resolution] graphs. Be sure to check off that option in **Options/Global settings...** of the **Graph** menu.

Terminology

We have called the **Plot tools and colors** *object-oriented.* What this means is that any shape, line, or text that you add is an *object* that you can move, remove, or change as a distinct entity.

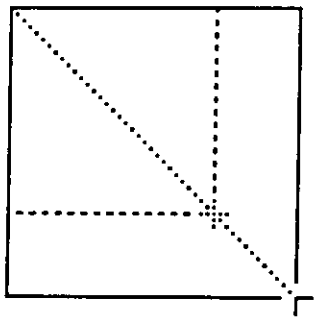
Note that the graph that SYGRAPH draws is also one object, and if you checked on the [Graphs scales as editable text] box in **Options/Global settings...**, each scale (the numbers showing the range of axis values) are separate objects also. You can draw, resize, and remove the objects SYGRAPH draws the same as you can the objects you add with the plot tools.

The only exception is that you cannot change the colors of the graph itself with the tools—you must choose colors with **Options/Color...** or the [Colors] option before plotting. You *can* change scale colors with the tools.

© 1989, SYSTAT, Inc.

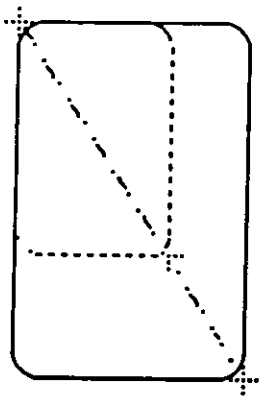
Rectangle

To draw a rectangle in the View window, click the rectangle icon in the toolbox. Position the pointer in one corner of the area where you want the rectangle, press the mouse button, drag to the diagonally opposite corner of the rectangle, and release the button.



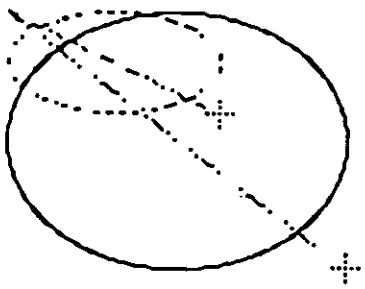
Rounded-corner rectangle

The rounded-corner rectangle works the same way as the rectangle. Click in one corner and drag to the diagonally opposite corner.



Oval

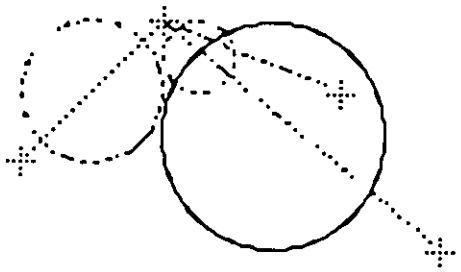
The oval works much like the rectangle and the square, except that the rectangular marquee you drag *circumscribes* the oval. Press the mouse button in one corner of the rectangle circumscribing the oval you want, drag to the diagonally opposite corner, and release the mouse button.



© 1989, SYSTAT, Inc.

Circle

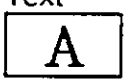
The circle works similarly to the oval, except that SYSTAT does not inscribe the rectangular marquee you drag. Instead, it draws the circle that has the *height* (vertical diameter) of the rectangular marquee you drag.



Line

To draw a line, click the line tool. Position the pointer at one endpoint of the line you want to draw, press the mouse button, drag to the other endpoint, and release the button. To add another line segment, press the button again at the intersecting endpoint, drag to the next endpoint, and release.

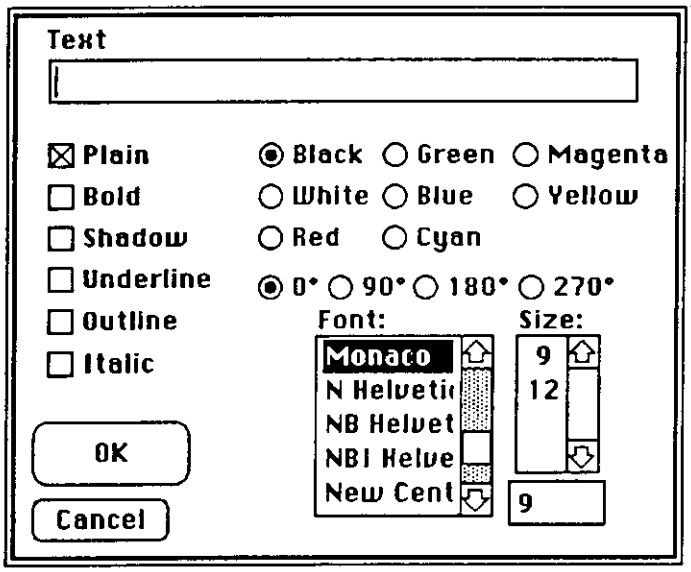
Text



To add text to a graph, click the text tool. Click at the spot in the graph where you want the text to start (the left end for horizontal text, the top end for vertical text). A dialog box lets you specify text and font.

- Type the text in the [Text] box
- Choose a color
- Choose a style
- Choose an angle for placement
- Choose a font and size

© 1989, SYSTAT, Inc.



To change the content, font, size, or style of any text already present, double-click the text object with the pointer. The text dialog box appears and lets you make your changes.

Scale labels can be altered with plot tools if you check the [Graph scales as editable text] box in **Options/Global settings...** from the **Graph** menu.

Pointer

You can move graph objects with the pointer. First, click any object to select it. Then you can move the object, resize it, remove it, or change its color.

Moving objects

To move the object, position the pointer inside the object, press the mouse button, drag the object to another location, and release the mouse button.

Resizing objects

Notice that when you select any non-text object, you see a gray "resize box" in the lower right corner of the marquee. Drag the resize box to change the size (and shape) of the object. Resizing with the gray resize box is illustrated in the first section of this chapter.

Removing objects

To remove a selected object, select **Clear** from the **Edit** menu.

© 1989, SYSTAT, Inc.

Colors

To change the color of any object, select a color from the toolbox (with the pointer tool) and then click inside the object. Recall that to control colors of the graph itself, you must use **Options/Color...** and/or the [Colors] option in the graph's dialog box.

Note that you cannot change the color of a graph with the plot tools. You must use **Options/Color...** or the [Colors] option in the graph's dialog box before plotting. You may change the color of graph scales if you had checked [Graph scales as editable text] in the **Options/Global settings...** before plotting.

Hints

If you have several objects on top of each other (or inside each other, or overlapping), you may have trouble selecting the one you want. This happens because your click inside one object may be interpreted as a click inside another object "on top" of it.

Think of objects as solid and layered. When you click at a given point, you will get the most "recent" object (i.e. the one that was added last) that contains that point.

To get at the objects "underneath" more recent objects, try clicking in different areas of the object. If you still have trouble, you may need to move the more recent objects temporarily out of the way.

© 1989, SYSTAT, Inc.

© 1989, SYSTAT, Inc.

9 Exploratory data analysis

Straightening scatterplots 166
SPLOMs 166
Straightening relationships 166
Outliers 170

Spinning 171
Spin 171
Home 172
X, Y, Z 173
Degrees 173
Labels 173
Background 173
Eye 174
Instant graphs 174
Brushing tools 174
Tutorial 174

Scatterplot brushing 182
Left arrow 183
Clicking on points 183
Clicking in white space 184
Marquee 184
Lasso 185
Subset analyses 185
Algebraic selection 186
Turning subset selections off 186
Extracting subsets 186

Tutorial 187

© 1989, SYSTAT, Inc.

Exploratory data analysis 9

Overview

The integration of SYSTAT and SYGRAPH into a single environment is an outstanding tool for exploratory data analysis, or EDA. A field pioneered by John Tukey, EDA is, in his words, "detective work—numerical detective work—or counting detective work—or graphical detective work." (Tukey, 1977).

In EDA, one examines data, particularly through graphing, to uncover as much as possible about the relationships lurking within. If one sees a direct relationship, one doesn't stop at a linear regression fit; one tries changing the scales to find a straighter plot. If one sees a particularly scattered plot, one smooths the value-to-value aberrations to see the general trends more clearly. If one sees a curved cluster of points, one tries logging and powering scales to straighten and narrow the cluster.

One works by trial and error, expecting that some trials will produce nothing helpful, but also expecting that one technique or another will elucidate the relationships among the data. SYGRAPH's graphing procedures, scatterplot brushing, data extraction, and smoothing offer much to the explorer.

This chapter assumes basic familiarity with the program. Please refer to earlier chapters and the *SYSTAT* and *SYGRAPH* volumes as necessary to refresh your memory on the mechanics of the program. The indices for each volume may be helpful. Instruction for two *new* techniques, 3D spinning and scatterplot brushing, *are* presented in this chapter.

The chapter concludes with an EDA tutorial.

© 1989, SYSTAT, Inc.

SPLOMs

In most situations, your first plan of attack should be to look at a SPLOM of all your variables. Looking at a SPLOM is a quick way to see what relationships might exist among your variables, because you see a tiny scatterplot of every possible pairing of the variables you select. For exploratory purposes, you may prefer a full SPLOM (with both upper- and lower-diagonal plots), since you get the chance to see each variable as both dependent and independent. Even though the plots on either side of the diagonal differ only by transposition, seeing both versions might encourage visual conclusions.

Look at the shapes of the point clouds in each cell. Are the clouds football-ish? Are the points tightly or loosely clumped? Do they line up on rows or columns? Does it look like a wet dog shook itself off and left a random splatter?

At the same time, consider your data. Do your variables record stimuli and responses? Different characteristics of a group? Survey results? Frequency distributions or grouped distributions?

If your predictor variables are *categorical* (grouping variables), you should have points stacking up horizontally or vertically in several lines. You may also think of such scatterplots as being like plots of continuous variables with slices removed. If both variables are *continuous*, the points will be scattered throughout the space more haphazardly. In either case—continuous or categorical—the same principles in searching for relationships apply.

Straightening relationships

To support any directional hypothesis, you ought to be able to fit a *regression* model to the data. In other words, if you think Y is some function of X, you ought to be able to get the points to straggle along a line of some sort.

For any of this to be relevant, your data must be continuous. If the variables are categorical (grouping variables), perhaps you should be comparing statistics from group to group.

© 1989, SYSTAT, Inc.

The simplest relationship looks like a straight line in a scatterplot. For instance, if Y is always 4X, the points line up. The difference between successive values (if sorted from smallest to largest) is *additive*; i.e. if you add some value to the first point, you get the second point, and adding the value again should give the third point, and so on. The equation of the line is $Y = mX + b$, where m is the slope of the line (rise over run), and b is the y-intercept (the value where the line crosses the y-axis).

If the line slopes the other way, the slope m is negative and the series is *subtractive*; e.g. subtracting something from the first point gives the second, etc.

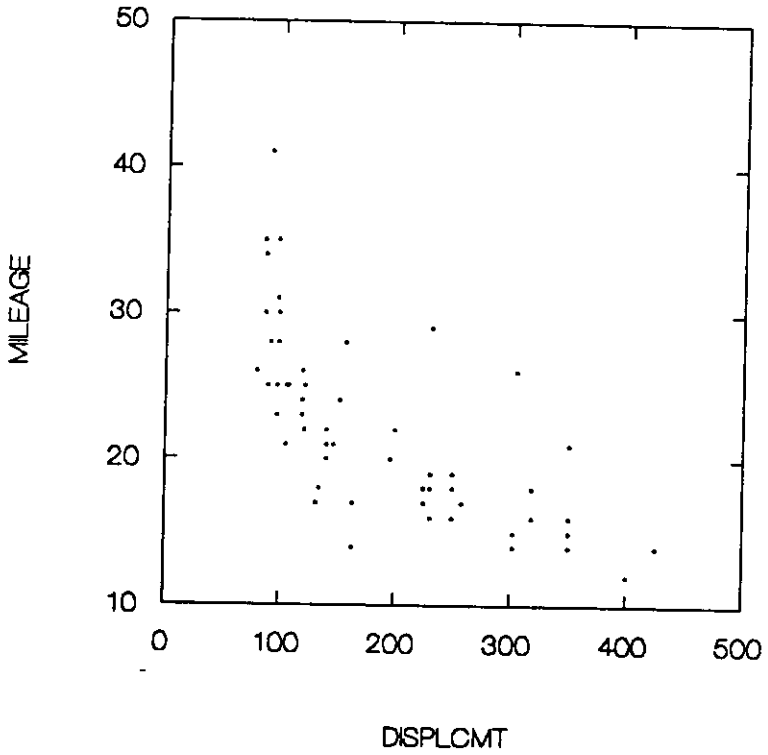
Both additive and subtractive slopes are *linear*.

Linear equations have attractive features for statistical modeling, so if the data are curved, you generally should try to see whether they can be made linear before using ordinary regression. And remember, using linear regression on untransformed curved data is absolutely invalid.

If the line is a curve, the series may be multiplicative, where one term is a given factor times the previous term. Try logging the variable. Since logarithms reduce multiplication to addition (or division to subtraction), replacing the values with their logs changes the curve to a line. Then, you can express a direct (linear) relationship of one variable to the log of another.

© 1989, SYSTAT, Inc.

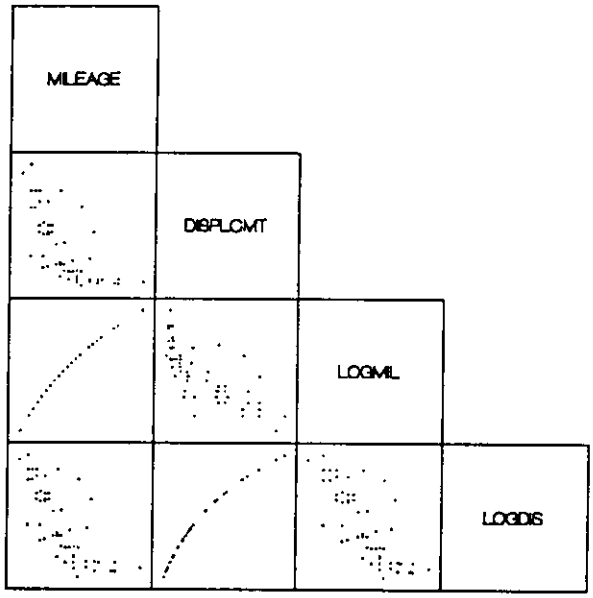
For example, a plot of MILEAGE against DISPLCMT (from the CARS dataset) has a distinctive curvilinear relationship:



Which variable should you log? While it is preferable to know the answer before you look at your data, there is nothing wrong with trying transformations until the data look linear.

First, create two new variables called LOGMIL and LOGDIS, which are the base ten logarithms of MILEAGE and DISPLCMT, respectively. You can do this with **Math...** from the **Editor** menu by setting LOGMIL to LOG(MILEAGE)/LOG(10) and LOGDIS to LOG(DISTANCE)/LOG(10). Then do a SPLOM of all four variables:

© 1989, SYSTAT, Inc.



MILEAGE versus LOGDIS remains curved while LOGMIL versus DISPLCMT is only slightly better. LOGMIL versus LOGDIS shows a trace of the original curve but clearly provides the most linear relationship.

You might also try some of the data transformations offered by **Series/Transform...**. Again, think about the data and make some educated guesses. Logging reduces multiplicative change to additive. Demeaning removes height: it moves the whole cloud down to center around y=0. "Square" squares each value (or undoes square roots). "Trend," "Percent Change," and "Difference" remove slopes, thus "untilting" clusters.

Logging is not the only transformation to consider for making curved relationships linear. Tukey and others have shown that "unbending" transformations form a hierarchy:

$$\ldots, x^3, x^2, x, x^{.5}, \log(x), x^{-.5}, x^{-1}, x^{-2}, x^{-3}, \ldots$$

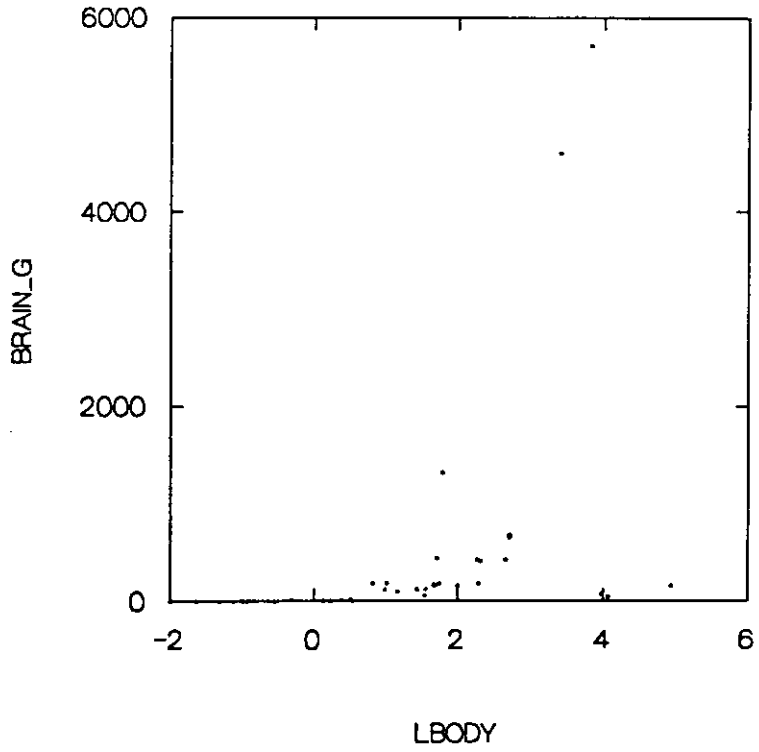
You may want to try some of these if logging doesn't unbend enough or if it unbends too much. For example, you can examine how inverting MILEAGE (creating a new variable GALLONS that is 1/MILEAGE) removes even more curvature than logging MILEAGE.

© 1989, SYSTAT, Inc.

Outliers

Do you have stray points even after, possibly, transforming your data?

If you have extreme outliers (points that don't seem to fit in), they're going to distort the scale and thus confuse the picture. Here is a plot from the BRAIN dataset of the base 10 log of brain weights versus the base 10 log of body weights. Notice that the two points near the top of the plot stretch out the scale and hinder us from examining the rest of the data.

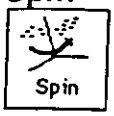


If you see a relationship that might make sense but a few points distort the picture, remove the points and look at what remains. This way, you can zoom in on more detailed variation.

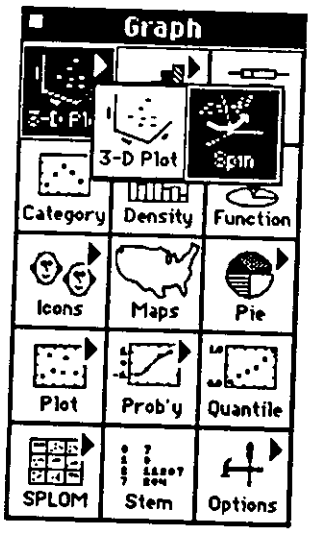
To remove outliers or look at subsets of the data, you might want to use SYSTAT's scatterplot brushing and extracting techniques, discussed below.

© 1989, SYSTAT, Inc.

Spin



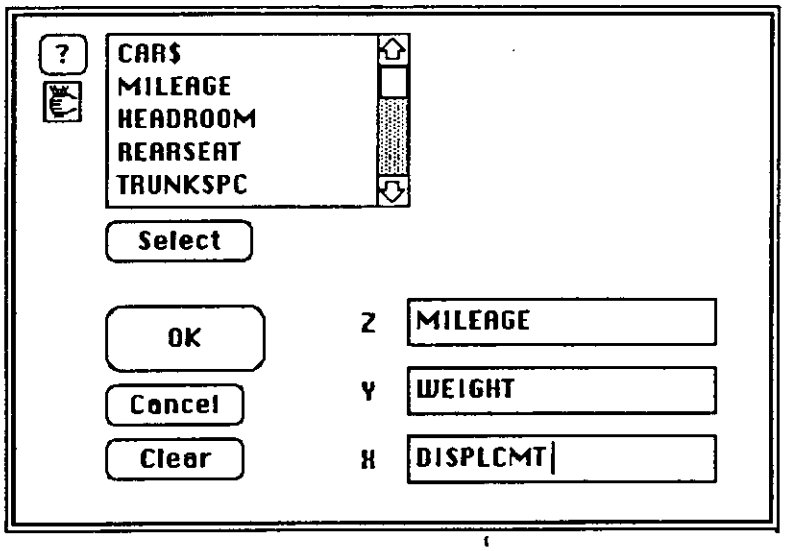
3-D Plot/Spin from the **Graph** menu lets you rotate three-dimensional scatterplots on all three axes in real time. Spinning plots in three dimensions often gives you a chance to see your point cloud from underneath, from behind, from the side—from literally any angle. Such an opportunity can often give you insights into your data that you would never get from seeing just one perspective plot.



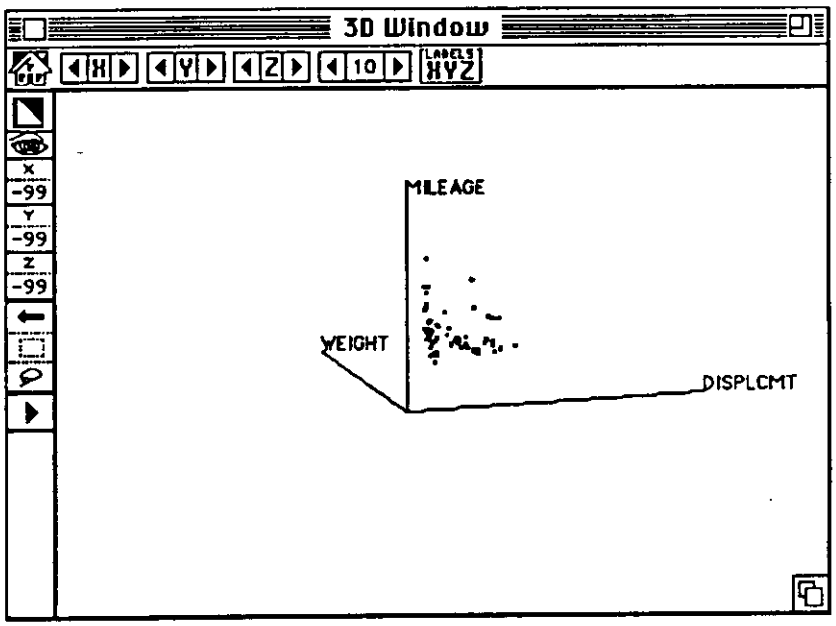
We will use the CARS dataset to show how **Spin** works.

- Select **3-D Plot/Spin** from the **Graph** menu
- Select the Z variable: MILEAGE
- Select the Y variable: WEIGHT
- Select the X variable: DISPLCMT
- Click OK

© 1989, SYSTAT, Inc.



SYGRAPH draws the plot in a special 3D Window.



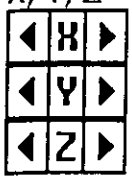
The 3D Window has a number of tools along the top and side.

Home

Clicking the Home button returns the plot to its original orientation, such as you see in the illustration above. You can click back home as often as you like.

© 1989, SYSTAT, Inc.

X, Y, Z

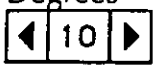


You can spin the plot on any of the three axes by pressing the arrow buttons on either side of the X, Y, or Z buttons. For as long as you hold the mouse button while pointing at an arrow, the graph rotates on the corresponding axis. To rotate in the other direction, use the opposite arrow.

You can switch from axis to axis or from one direction to another as much as you want.

By clicking the middle portion of the button (the actual X, Y, or Z), you can remove the corresponding axis line from the plot.

Degrees



Use the degrees control to change the rate at which the plot spins. Notice that the spinning is accomplished by animation: the plot is shifted from one position to the next quickly, creating the illusion of motion. The default is for the plot to shift ten degrees at a time, but you can make it spin more smoothly and slowly by decreasing the shift, or faster and jerkier by increasing the shift.

To decrease the amount of shift, press the leftward arrow button; to increase, press the rightward arrow. You can set the shift amount to any even degree value between 2° and 90°.

Labels

Notice that the three axes are labeled with their variables, MILEAGE, WEIGHT, and DISPLCMT. You can remove any or all labels from the plot by clicking the X, Y, or Z of the [Labels] option. If you click a label off, its letter on the option is dimmed; to restore the label, click it back on.

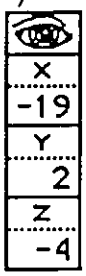
Background

By default, **Spin** draws a black plot against a white screen. If you prefer, you may have a white plot against a black screen. Just click the black/white button to switch back and forth.

If you use **Options/Color...** from the **Graph** menu to set background, foreground, graph, and label colors, **Spin** plots with the colors you set.

© 1989, SYSTAT, Inc.

Eye

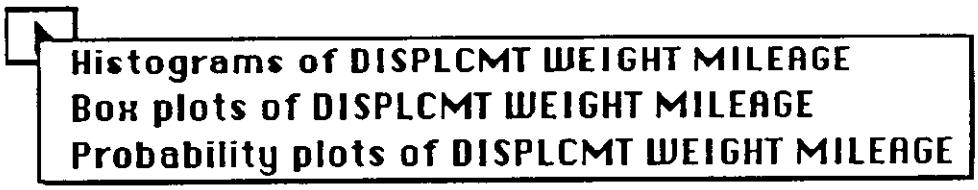


After you have spun the plot and found what you consider to be the optimum perspective, click the [Eye] button. The X, Y, and Z boxes show what settings to use for the [Eye] option in **Options/Formats...** so that you can print a final, high resolution **3-D Plot** for presentation.

If you, like many people, find the [Eye] option in **Options/Formats...** confusing, just stick with **Spin**. After you've found the orientation you like, check the [Eye] coordinates, set them, and then whip out a final copy with **3-D Plot**.

Instant graphs

The arrow button at the far right leads to a pop-up menu providing instant histograms, box plots, or probability plots. After spinning a plot of three variables, you might want to see quick histograms—or box plots—or probability plots—of each variable. Rather than having to go to the **Graph** menu and play with a bunch of dialog boxes, you can just use the pop-up menu. Point at the arrow and press the mouse button. A little menu pops up; select the item you want and release the mouse button.


Histograms of DISPLCMT WEIGHT MILEAGE
Box plots of DISPLCMT WEIGHT MILEAGE
Probability plots of DISPLCMT WEIGHT MILEAGE

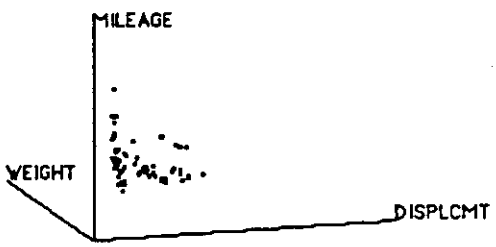
SYGRAPH does the rest. All you do is decide whether to stack or overlay the graphs in the View window.

Brushing tools

There are three scatterplot brushing tools available if the Data Editor is open when you use **Spin**. These tools work the same as the brushing tools in the regular View window. See the "Scatterplot brushing" section, next, for details.

Tutorial

Above we showed the original orientation of our MILEAGE by WEIGHT by DISPLCMT plot. No, you don't have to flip back to it. Here it is again:

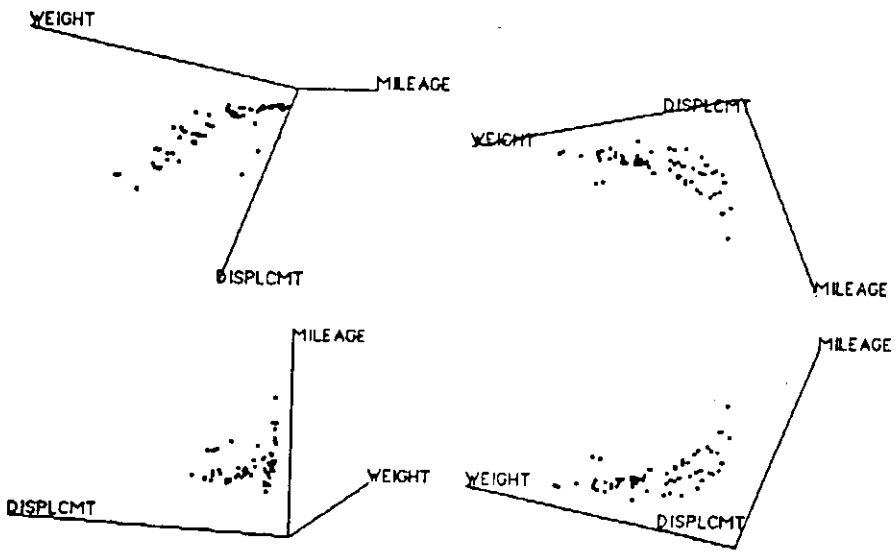


© 1989, SYSTAT, Inc.

We often orient plots a certain way to suggest causality: in an X-Y scatterplot, we plot the predictor as X and the response as Y. In an X-Y-Z scatterplot, we might have two predictors as X and Y and a response variable as Z. This custom often leads us to infer a causal relationship when we see a 3-D plot. We look for a way to interpret the vertical variable as some function of the "floor" variables, X and Y.

In fact, the axis assignments are as likely as not to be arbitrary. This example is a good case in point. Most of us know that fuel efficiency, vehicle weight, and engine displacement are all related characteristics, so we chose those variables for the example. We then asked you to plot those three variables against each other. Since we didn't have any pet theory to bang you over the head with, we suggested an order arbitrarily.

So let's question the axis assignments right away. Use the degrees control to set the shift amount up to 90°. Now, click any of the X, Y, or Z buttons by turns to flip the plot. This amounts to switching axis assignments. Here are some flips:



You can try some of your own. In the upper right plot notice that the DISPLCMT axis is invisible; this is because it is approximately perpendicular to the "plane" of the monitor screen or the page of this book. When you're spinning the graph, you can see more easily whether the axis is pointed toward or away from you.

© 1989, SYSTAT, Inc.

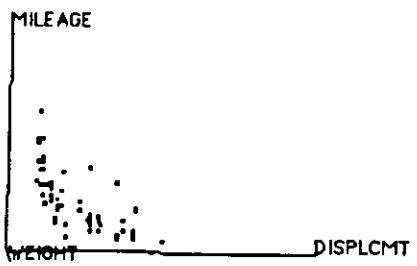
You've probably already noticed something else. What used to look like a tight little glob of points floating in space is clearly not. In a few of the plots above, the point cloud seems to be spreading across a sort of curved surface. We can try to see this more clearly by spinning along one axis at a time in smaller increments.

So, use the degrees control to set the shift back down to something small—say, 6°. To make it easier for you to follow along, we'll also click Home to start fresh again.

Now, press the mouse button on the leftward X arrow. What we're going to do is make MILEAGE perpendicular to the screen so we can look at WEIGHT and DISPLCMT. As it spins, you can see the following come into view:

We see an almost linear relationship between WEIGHT and DISPLCMT. That makes sense: bigger cars have bigger engines. We might want to try some transformations later to see if we can straighten it up a bit. Notice that the points fan out somewhat as they go away from the vertex of the three axes. We might also try to bring that under control.

Now, let's look at MILEAGE and DISPLCMT. Click Home again, and aim for this orientation. We used the leftward Z, the leftward X, and the leftward Y axes each briefly, in that order; you could get a similar orientation with a different sequence of moves.



© 1989, SYSTAT, Inc.

Anyway, MILEAGE and DISPLCMT are not so easy to interpret. First, the clump is "upside-down;" we generally like to see relationships that are more like Y=X. We could fix that problem by using the reciprocal of MILEAGE, GALLONS, like we did earlier.

Also, notice that the points are scattered more randomly than we might hope. We might try transforming the variables to find a more linear shape.

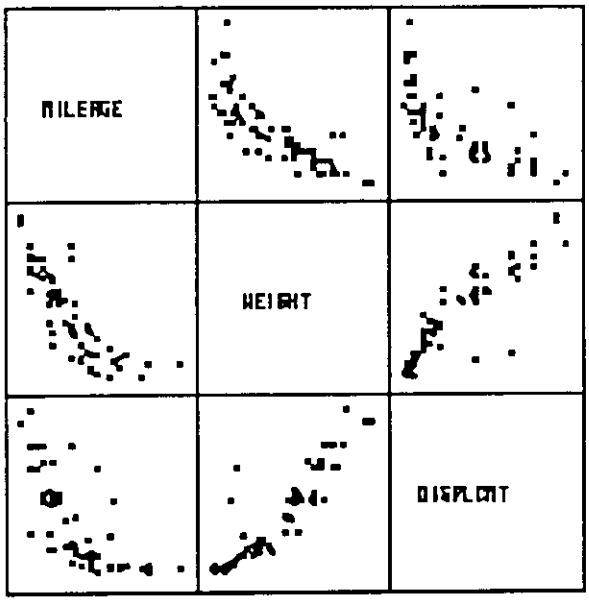
Again, what we see makes sense. Larger engines (which we might casually associate with bigger cars) generally are less fuel efficient, but there are big cars that are efficient and little cars that are not. Other factors are involved: actual car size (WEIGHT), number of cylinders, type of transmission, and so on.

Finally, let's look at MILEAGE and WEIGHT. We'll let you figure out how to spin to an orientation like the one below.

We see the same problem as before: we'd rather have the inverse of MILEAGE.

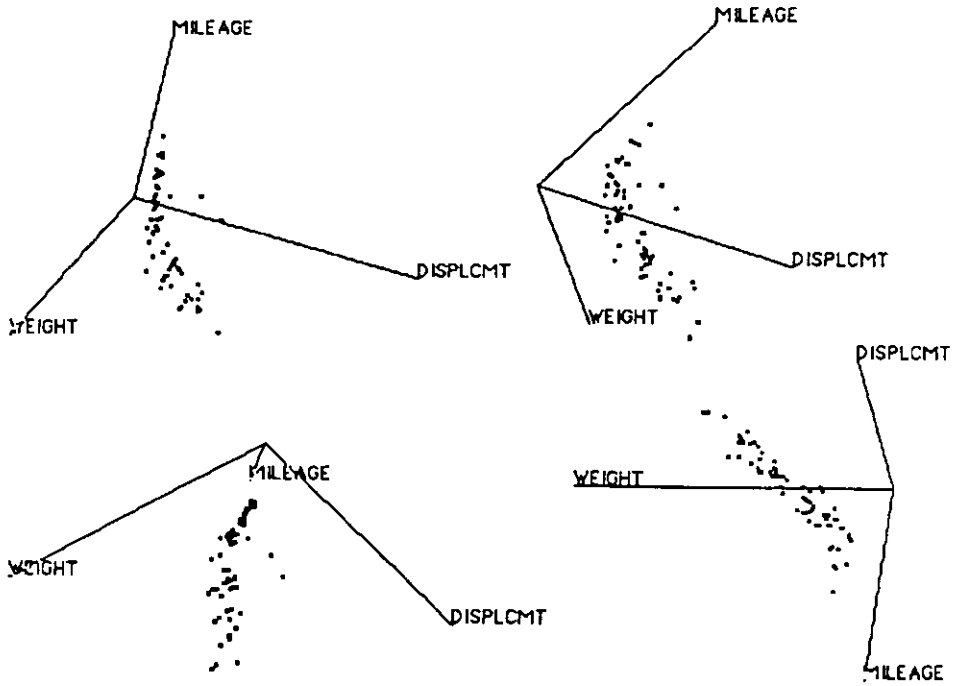
© 1989, SYSTAT, Inc.

By now you're probably wondering why we didn't just look at a SPLOM if we wanted to see each pairing. Good question. Let's do a SPLOM now. Since we're exploring, we'll draw a full symmetric SPLOM.



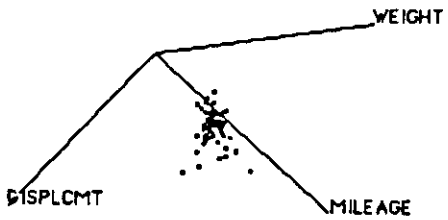
This is unquestionably more efficient. But what we don't see in the SPLOM are angles "above" and "below" and "beside" the cloud of points. You know a shape better when you see it from all sides. When you were spinning, perhaps you noticed some views like these:

© 1989, SYSTAT, Inc.



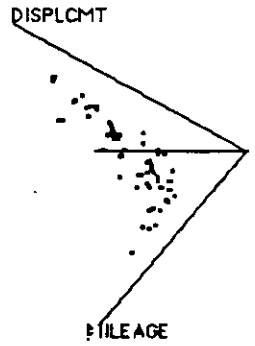
Another thing we don't see in the SPLOM is the motion. Sometimes it is the spinning process itself that is most enlightening: by seeing a two-dimensional representation of three dimensions "in motion" we are better able to construct a three-dimensional image cognitively. (It is easier for most people to perceive depth from a motion picture than from a snapshot.)

If you want to mislead your audience, you might find a view like this useful:



© 1989, SYSTAT, Inc.

Here, we clicked off one axis label that was in the way:



Here are some views without the axes. Removing the axes might free you to visualize some other axes, such as might be produced by principal component analysis. (In fact, you might want to try spinning your **Factor** results.)

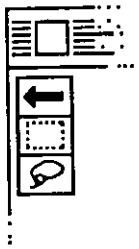
© 1989, SYSTAT, Inc.

You might look for the orientations that maximize or minimize spread along one (computed or visualized) component.

© 1989, SYSTAT, Inc.

Scatterplot brushing

Scatterplot brushing allows you to isolate data points visually. These scatterplot brushing tools appear in the upper left corner of the View window any time you draw a scatterplot or a SPLOM with the Data Editor open.



You can use the left arrow to select a point and see which case generated it by looking in the Data Editor, or to select one or more cases in the Data Editor and look in the plot to see the point(s). You can use the lasso or marquee to select a group of points. You can then do analyses on just those cases, or save those cases in a new file for analysis later.

This section gives instruction for each of the tools. The final section, "Tutorial" is a walk-through demonstration of how to use scatterplot brushing on the BRAIN dataset that compares brain weights and body weights of a variety of animals (from tiny rodents to huge dinosaurs).

Problems?

I don't see the brushing tools.

1) Be sure to open the file in the Data Editor. The brushing tools only work when the Data Editor is open.

2) The brushing tools disappear whenever you delete a row or case from the Data Editor, change or delete a data point, resize the View window, or resize or relocate the graph.

© 1989, SYSTAT, Inc.

I can't use the resize box or move the graph. You need the regular pointer cursor to move or resize graphs. Just click off the currently selected tool, and you have the pointer. Remember, though, if you resize or move the graph, the brushing tools disappear. Position and size the graph *before* plotting with the ruler option.

Left arrow

The left arrow has two functions. First, you can click on a point in a plot or SPLOM to see which case generated it and, in the case of a SPLOM, which points in the other plots of the SPLOM were generated by the same case. Second, you can click in white space of the View window to highlight the points generated by the cases that are selected in the Data Editor.

Clicking on points

You can click a point in the plot to highlight its case in the Data Editor. If you have a SPLOM, the points generated by that case in other plots of the SPLOM are highlighted for as long as you hold the mouse button.

- Be sure the View window is active
- Choose the left arrow tool
- Click on a point

Now, look in the Data Editor. The case that generated the point is highlighted.

You may want to move the View and Editor windows so that you can see the Data Editor clearly even when the View window is the active window.

Problems?

I don't see any highlighted cases.

1) Be sure that you clicked *on* a point and not near it.

2) If you have several points clustered closely together, you may have to try a couple times to get the one you want. If it is too difficult to isolate a point, try selecting a group (discussed below) or adjusting [X min], [X max], etc. with the [Axes] option and replotting to "zoom in" on the area you're interested in.

3) If the point came from a case that is not currently visible in the Editor window, you may need to scroll up or down.

© 1989, SYSTAT, Inc.

4) If you had selected any cells in the Editor before using the left arrow, the case is highlighted in that column—the "current" column. Decide which variable column you would like the left arrow to highlight, click somewhere in the column to make it the "current" column, and try again.

Clicking in white space

The second function of the left arrow is to see in the plot what cases are currently highlighted or bulleted.

- Highlight a case by using the left arrow, bullet points with the marquee or lasso, or select cases in the Data Editor
- Choose the left arrow tool
- Press the mouse button anywhere in the View window *except* on a point—that is, click on white space
- When you are done studying the highlighted points, release the mouse button

If you want to see the points for a longer time but don't want to hold onto the mouse button, hold the Option key and click the left arrow. The points will blink on and off until you click again.

Problems?

I selected a bunch of points, but only one point is highlighted. Be sure that you are clicking on white space. If you click on a point, only that point (or that point in all plots of a SPLOM) is highlighted. Click outside the plot to be sure.

The wrong points (or no points) are highlighted. As we discuss below, you "deselect" points by clicking the lasso or marquee in the View window. If you clicked in the window before choosing the left arrow tool, you may inadvertently have turned off or changed the selection. Also, be sure you are not clicking the left arrow on a point.

Marquee

Use the marquee to make rectangular selections of points. Any points you select in the plot are marked with round bullets (•) in the Case column of the Data Editor.

- Choose the marquee tool
- Press the mouse button and drag a rectangular marquee around the points you want
- Release the mouse button
- Look in the Data Editor to see the bulleted cases

© 1989, SYSTAT, Inc.

You can use the Shift key to make cumulative selections. Make one selection the usual way, then hold the Shift key while making another. Both sets of points are highlighted.

Once cases are bulleted in the Editor, any procedures you perform are limited to those cases. So, you can use the marquee to select a subset of cases, do **Stats** and **Graph** analyses of the subset, and then return to the full dataset. More on this under "Subset analyses."

To turn off a selection (bulleting), just click the marquee anywhere in the View window.

Lasso

Use the lasso to rope irregularly-shaped clusters of data points. Any points you select in the plot are marked with round bullets (•) in the Case column of the Data Editor.

- Choose the lasso tool
- Press the mouse button and draw a rope around the points you want
- Release the mouse button
- Look in the Data Editor to see the bulleted cases

If you want to isolate several remote clusters, either draw a tricky rope that doesn't capture extra points or make cumulative selections.

To make cumulative selections, make one selection the usual way, then hold the Shift key while making another. Both sets of points are highlighted.

Once cases are bulleted in the Editor, any procedures you perform are limited to those cases. So, you can use the marquee to select a subset of cases, do **Stats** and **Graph** analyses of the subset, and then return to the full dataset. More on this under "Subset analyses."

To turn off a selection (bulleting), just click the marquee anywhere in the View window.

Subset analyses

Once cases are bulleted in the Data Editor, you can do further analyses with just those cases. For instance, you could use **Corr** to find the Pearson correlation coefficients of *just those cases*; thus, you could remove the effects of outliers on the correlation of an otherwise convincing cluster.

© 1989, SYSTAT, Inc.

Or, you could use **Plot** again and see just those points more closely: since you reduce the range of values, the scale becomes smaller. You're effectively "zooming in" on that part of the graph. You could then do a further selection by using the marquee or lasso on *that* plot to narrow the field still more.

Algebraic selection

Editor/Select... lets you to select cases algebraically. This is discussed in the **Data Editor** chapter.

Turning subset selections off

Any bullet selection remains in effect until you make a different selection, click the marquee of lasso in the View window to turn the selection off, select **Remove select** from the **Editor** menu, open a new file, transform a case, enter a new value in the Editor, or delete a case.

Extracting subsets

You can extract bulleted cases into a separate file with **Extract...** from the **Editor** menu.

- Use the marquee, lasso, or **Editor/Select...** to select cases
- Select **Extract...** from the **Editor** menu
- Specify an output filename and location
- Click Save

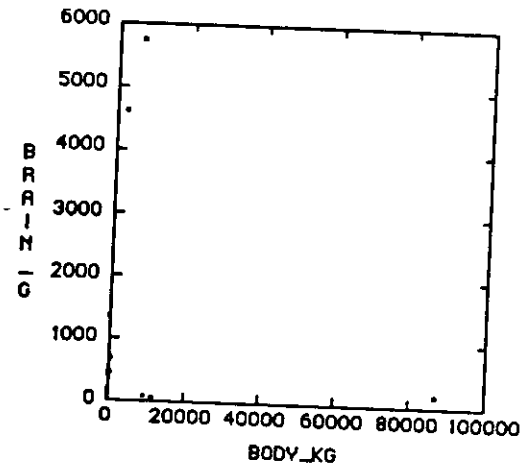
All the bulleted cases are placed in the new file. You can then open that file and do further analyses on just those cases.

© 1989, SYSTAT, Inc.

Perhaps the best way to introduce scatterplot brushing is with an example. The BRAIN file is a sample dataset recording the body weights (in kilograms) and brain weights (in grams) of various species.

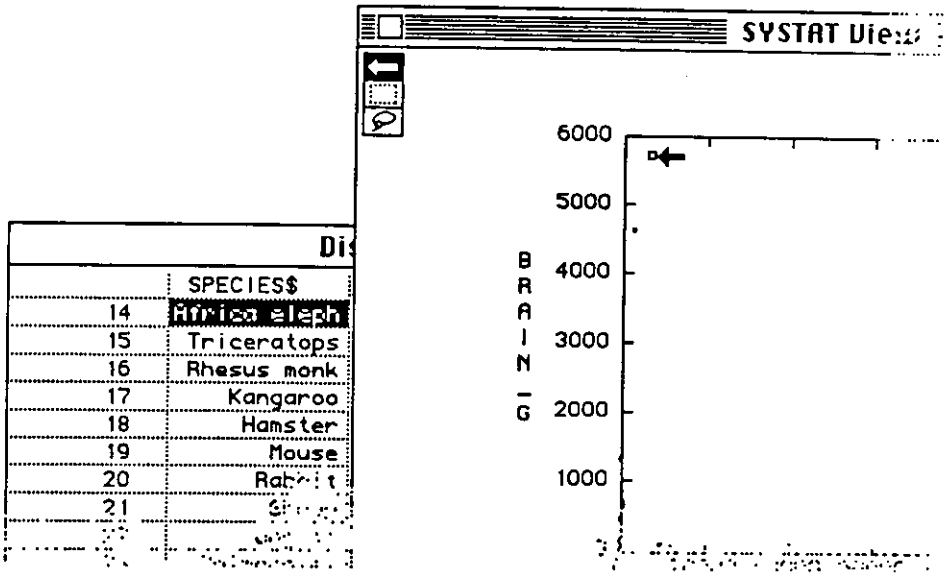
Before we start, we know certain things. Dinosaurs and elephants are big. Very big. Rats are not. But are dinosaur brains as large, proportionally, as elephant brains? Are rat brains as small as they would be if rats were tiny dinosaurs?

Let's look at a scatterplot of BRAIN_G against BODY_KG:



© 1989, SYSTAT, Inc.

Right away, we know a few things. First, we have three points that are way out of range. Choose the left arrow tool and click the topmost outlier, then click on the Data Editor to make it active.

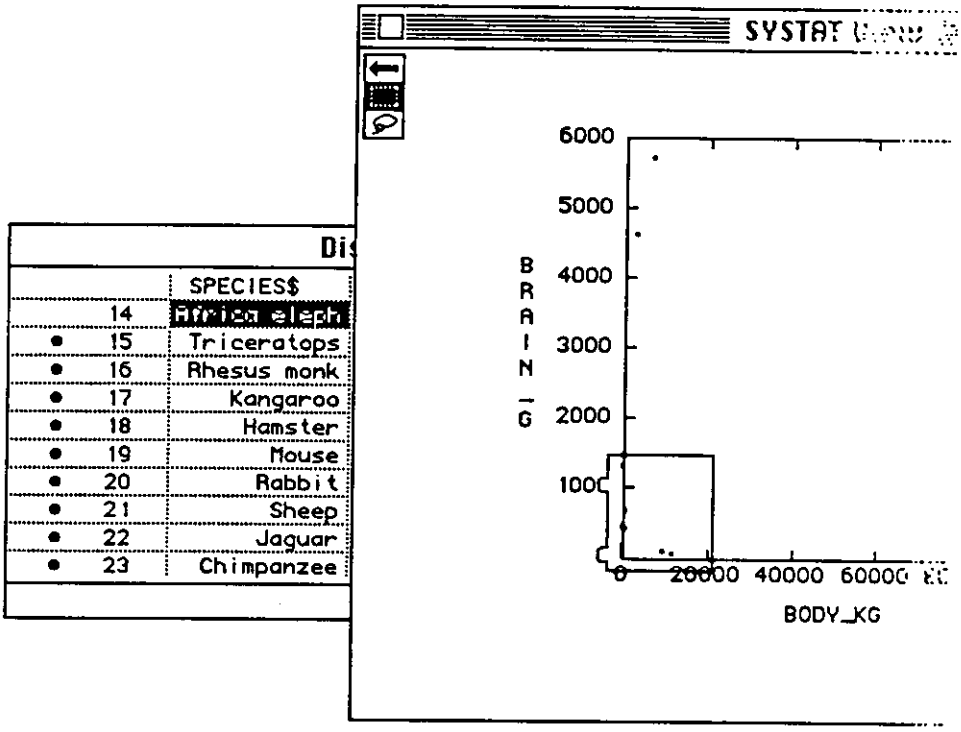


We see that this huge-brain critter is the African elephant. Clicking on the other two outliers reveals a huge-brained Asian elephant and a big, stupid Brachiosaurus. All three are out of range similar distances (visually), but note that elephants have disproportionately large brains but body sizes fairly well in range. The Brachiosaurus, contrarily, has a huge body but an embarrassingly small brain.

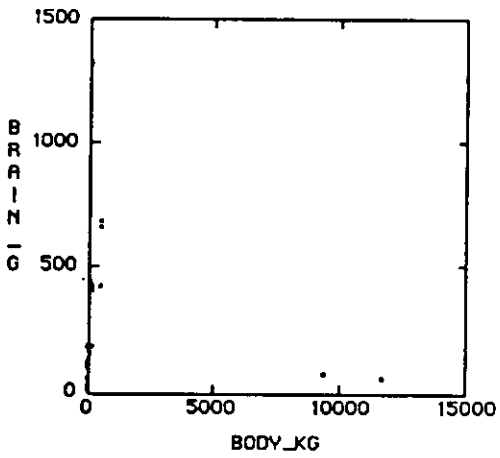
The remaining points represent animals whose brain/body weight ratios are closer to the average. However, notice that we see only about a dozen points on the plot, yet we have a whole bunch of cases. We must conclude that the lower corner of the graph has points stacked on top of each other.

Choose the marquee tool and pull a marquee around the lower clump. After a pause, the selected points blink and their cases are bulleted. Any analysis we do now will use only those bulleted cases.

© 1989, SYSTAT, Inc.



Use **Plot** again. (It might be easier to use **Redo last analysis**, abbreviated ⌘R.)



© 1989, SYSTAT, Inc.

We still have a tight concentration in the lower left corner and a bunch of brainy or overgrown outlying animals. Again using the left arrow tool, we see that humans, horses, giraffes, chimpanzees, donkeys, gorillas, and cows are disproportionately large brained. The latter five points are clumped closely together—you may need to fuss a little bit and point carefully to find out all the mammal types where the points stick together. Or, you could use the command key (⌘) and the marquee or the lasso, and then look at the cases that are bulleted for extraction.

By now, though, you should have made an important observation: the data are tightly scrunched together near the origin and fan out increasingly as BRAIN_G and BODY_KG increase. In fact, the increasing amount of space looks almost—could it be?—multiplicative in both directions!

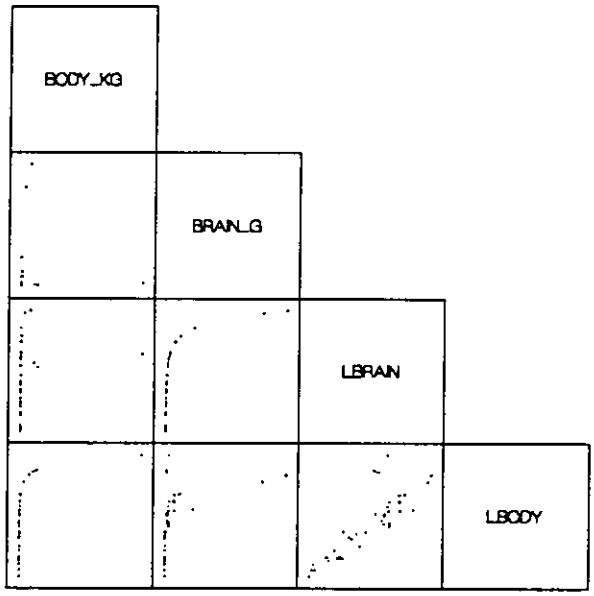
You should be thinking, "What if I take the logarithms of the variables?"

What, indeed? Try it. Use **Editor/Math...** to create new variables LBRAIN and LBODY:

```
Set LBRAIN to LOG(BRAIN_G)/LOG(10)
Set LBODY to LOG(BODY_KG)/LOG(10).
```

Now, look at a SPLOM of all four variables: BRAIN_G, BODY_KG, LBRAIN, and LBODY. Recall that a SPLOM is almost always the best bet when you're beginning any analysis, because it shows you *every* possible pairing of variables. If you didn't use a SPLOM, you might start with a plot of LBRAIN and BODY_KG. As you can see in the SPLOM, even logging just BRAIN_G cleans up the plot a lot, and you might be content to stop there. The SPLOM, however, makes it clear that logging both variables lends the clearest relationship.

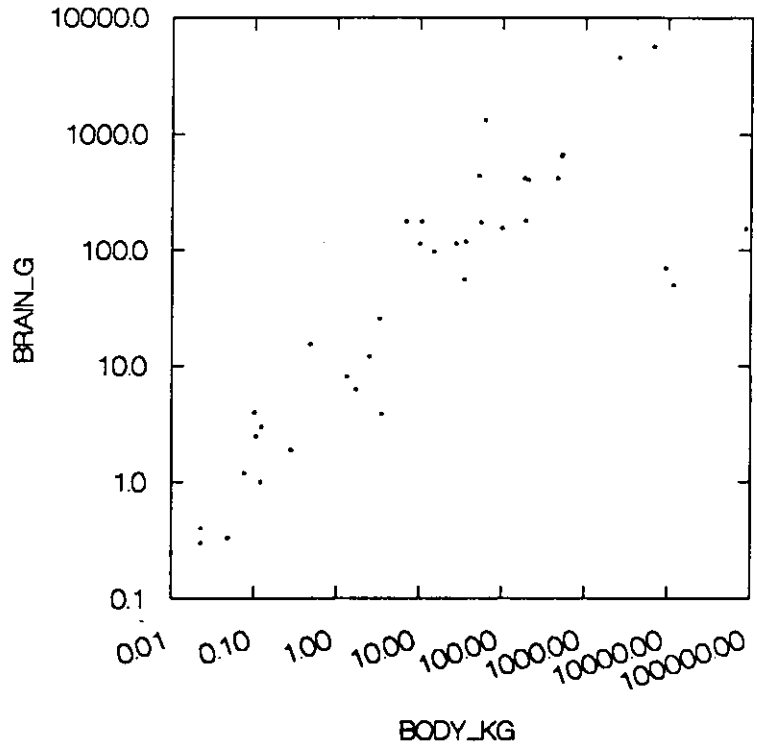
© 1989, SYSTAT, Inc.



Now we plot BRAIN_G against BODY_KG using the [LOG] option to log both variables automatically.

- Select **Plot/Plot** from the **Graph** menu
- Select the Y variable: BRAIN_G
- Select the X variable: BODY_KG
- Click [Log] and specify 10 for [X log] and [Y log]
- Click OK

© 1989, SYSTAT, Inc.



We see how scatterplot brushing and extracting can help us see relationships we might not have guessed otherwise. Once we extracted a few outliers, the plot got enough clarity that we could guess more easily that logging might be helpful.

At this point, you might want to continue with some more brushing and extracting. You'll notice that the dinosaurs are still outliers on the underside of the average brain-to-body proportion, although certainly not to the degree they were before logging the data.

Make your own decisions: for your purposes, is it valid to narrow the field to a few millennia worth of animals? If so, then extract away. Or, throw validity by the wayside for now, and make some experimental, arbitrary extractions and do some stats—see how your results change.

© 1989, SYSTAT, Inc.

© 1989, SYSTAT, Inc.

10 Tips and shortcuts

All the windows are text editors 195
Redo last analysis 195
Save dialog settings 196
Bundles 196
Help 196
Scatterplot brushing tools 196
Command-key abbreviations 197
Command interface 197
Command logs 197
Command batches 197
Instant graphics and statistics 197
Variable names 198
Matrix cells 198
Example 198

© 1989, SYSTAT, Inc.

Tips and shortcuts 10

All the windows are text editors

In the **Redirecting output** chapter, we explained that both the Analysis window and the Command window are text editors. You can type in them to add your comments. Just place the cursor where you want to type something, click the mouse button, and start typing. You can add new lines by pressing Return, and you can remove lines and material by selecting it with the mouse and pressing Delete.

(Note that if you are editing in the Command window, you have to be careful about prompts; see the **Command reference** appendices in the *SYSTAT* and *SYGRAPH* volumes for information about editing command batches.)

You can also **Copy** from them, **Cut** from them, and **Paste** into them. You can cut the numbers that some analysis spat out, paste them into the Data Editor, and analyze those results.

If you choose [Scroll analyses] from **Data/Formats...**, you can let all your analysis results pile up in the Analysis window until you're done working. Then, select all the contents of the window (use **Select all** from the **Edit** menu) and print them and/or save them. You'll have a permanent record of your analysis process; if you ever need to reconstruct your efforts, it will be easy.

Do you find yourself sticking little notes all over your machine reminding yourself to double check a data point or rerun some analysis without the outlier? Type those notes directly into the Analysis window, and they won't get lost.

Redo last analysis

You can redo your most recent statistical or graphics procedure in one step by selecting **Redo last analysis** from the **Goodies** menu. **Redo** repeats the most recent action taken through the **Stats** or **Graph** menu.

For example, after you bullet some points with the scatterplot brushing tools, you might want to look at a plot of just those cases. Select **Redo last analysis**, and SYGRAPH automatically draws a new plot, only this time it uses only the selected cases, just as it would if you had gone to the **Graph** menu yourself.

© 1989, SYSTAT, Inc.

Or, suppose you had done a SPLOM of all the variables in a file and then you decided you needed to log a few variables. After you get done with **Editor/Math...**, you're ready to plot again. Just select **Redo last analysis**.

Save dialog settings

In the **Analysis** chapter we showed how you can save all your dialog box settings with the [DLOG] option in the **Save as...** dialog box. Saving dialog boxes—particularly when you're working on a lot of repetitive models with **MGLH**—can save you a lot of pointing and clicking when you go back to work after lunch or the next day or a year later.

You might want to set up special DLOG folders—one for each dataset you work with—and include the date in each filename. You can keep a dialog box "journal."

Bundles

Also in the **Analysis** chapter, we discussed *variable bundles*. Whenever you're using a dataset with more than a dozen or so variables, you will probably notice that you use only a handful of variables at a time. You waste a lot of time scrolling through variable selection lists.

Set up a few bundles of your most frequently needed handfuls of variables. Remember, you can switch from one bundle to another or back to the whole dataset instantly.

Help

If you forget how to run an analysis or produce a graph, you can probably find the information you need most quickly with the on–screen help facilities.

Don't forget that help is a click away for every dialog box, button, check box, selection list, window, statistical term, tool, command,... See the **Help** chapter for details.

You probably won't need to use these manuals for weeks at a time after you've gotten the hang of SYSTAT and SYGRAPH. (We hope you *will* read the manuals, of course—but you shouldn't have to go digging through indexes for instructions on a button!)

Scatterplot brushing tools

Need to identify an outlier? Extract a subset? Exclude some outliers? Follow some cases across the various plots of a SPLOM? Scatterplot brushing tools let you do these things quickly, visually.

© 1989, SYSTAT, Inc

Command-key abbreviations

Many of the menu items have command-key abbreviations. Rather than having to use your mouse to pull down the menu and select an item, you can type its command-key abbreviation. Command-key abbreviations use the Command (Apple or ⌘) key much like the Shift key: hold the Command key, press the appropriate letter, and release the Command key. Like most Macintosh applications, for instance, you can use Command-O (⌘-O) to open files, ⌘-S to save, etc.

Command-key abbreviations are shown in the menus along the right side.

Command interface

Many people find command interfaces to be more efficient than menu interfaces for some types of work. Particularly for repetitive analyses, SYSTAT and SYGRAPH's optional command interface can be helpful. See the introductions to the **Command references**, which appear as Appendix I in both the *SYSTAT* and *SYGRAPH* manuals, for more information about how and why to use commands.

Command logs

A particularly useful feature of the optional command interface is that it enables you to keep precise, complete records of every step you took in analyzing and graphing your data. Even if you actually use the menus to do your work, you can have the Command window open and SYSTAT and SYGRAPH will automatically generate a command log of your work. At the end of the session, all you need to do is select all the material in the command window and print it or save it. Again, see the **Command reference** appendices for details.

Command batches

You can also use the command interface to put together *batches* of commands to run all at once—perhaps while you go to lunch. You can put batches together easily by editing the command log of a prior session. See the **Command reference** appendices for details.

Instant graphics and statistics

After results of statistical procedures are displayed in the Analysis window, SYSTAT can suggest some possible next steps—other graphical or statistical displays that you might want to check—and even perform them for you. Just hold the Option key to get the magnifying glass cursor and press the mouse button on any variable name or any cell in a matrix.

© 1989, SYSTAT, Inc

Variable names

In any statistical output, if you hold the Option key and press the mouse button on a variable name, SYSTAT offers a pop-up menu of graphics and statistic that you might want to try.

This pop-up menu works like a regular Macintosh menu: select the item you want and release the mouse button. If you decide you don't want to do any of the suggestions, just move up and off the menu before releasing the mouse button.

If you do select an item, SYSTAT does all the work for you. You don't have to go to the **Graph** or **Stats** menus and click your way through a dialog box; you just decide whether to overlay or stack graphs.

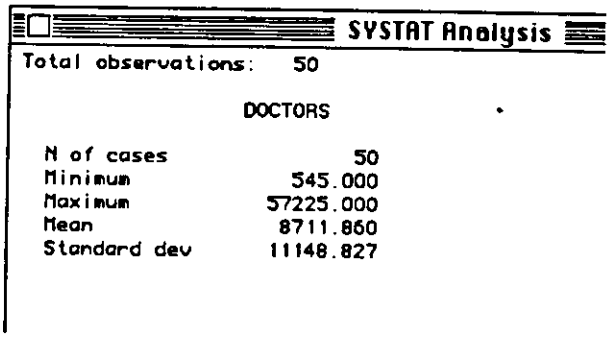
Matrix cells

In any matrix produced in the Analysis window (such as a matrix of Pearson correlation coefficients), you can Option-click any cell to get a menu of graphics and statistics for the corresponding pair of variables.

Again, this menu works like any menu, and SYSTAT does all the work.

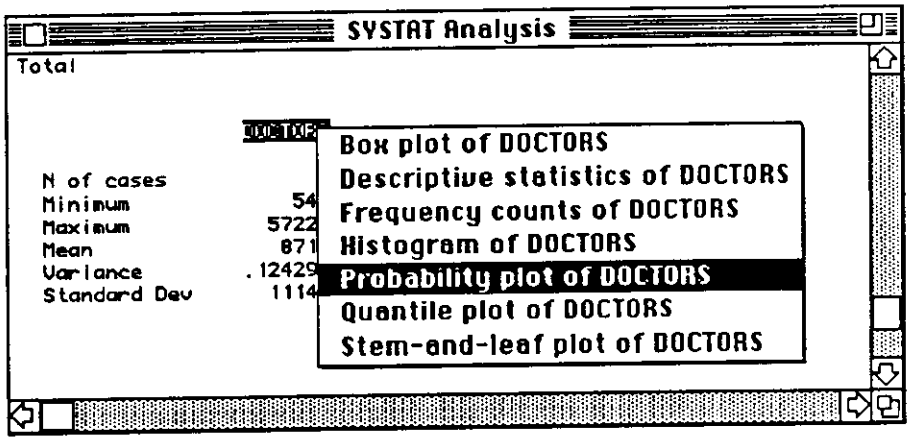
Example

Stats/Statistics... produces the following for the DOCTORS variable in the USDATA file:

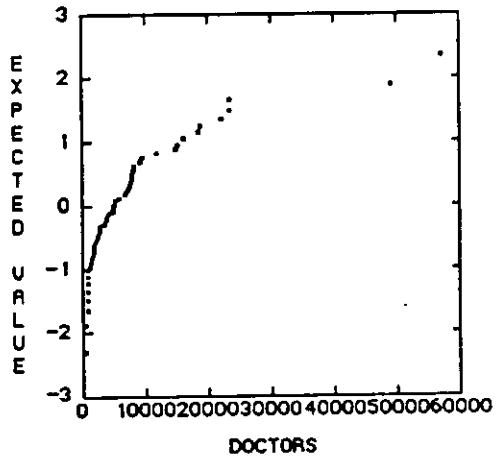


Now, if you hold the Option key, the cursor changes to a magnifying glass. With the magnifying glass, press the mouse button on the variable name DOCTORS. The following menu pops up:

© 1989, SYSTAT, Inc

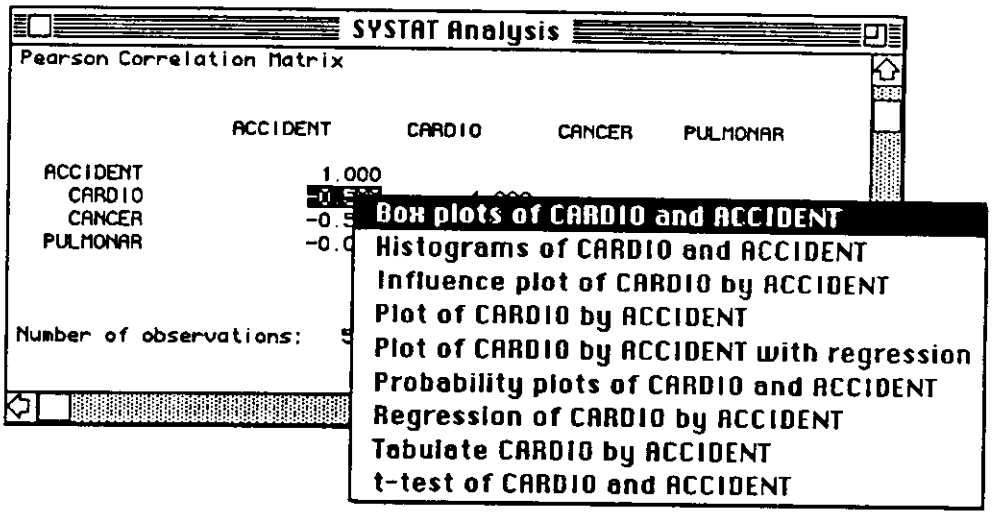


We release the button on the **Probability plot of DOCTORS** item and get the following plot:

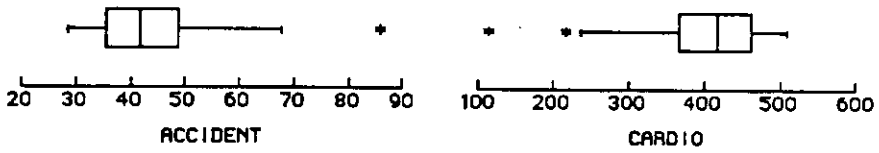


Next, we use **Corr/Pearson...** from the **Stats** menu to produce a correlation matrix of ACCIDENT, CARDIO, CANCER, and PULMONAR. We can Option-click the CARDIO, ACCIDENT cell and get a menu of suggested statistics and graphics for CARDIO and ACCIDENT:

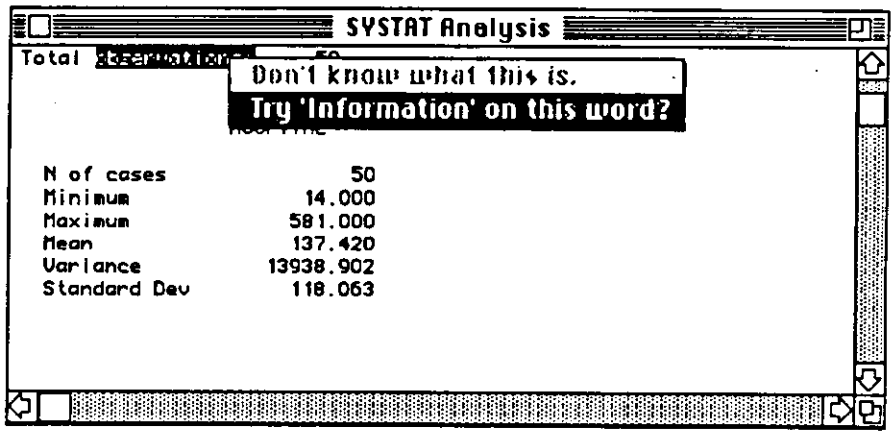
© 1989, SYSTAT, Inc



We select **Box plots of CARDIO and ACCIDENT** and get the following box plots, instantly. As with all graphs drawn in the View window, you must choose [Overlay plot], [Stack plot], or [Next plot] between graphs.



If you try to Option-click some other type of item in the Analysis window, SYSTAT reports that it doesn't know what it is and then offers to try **Information**.



© 1989, SYSTAT, Inc

Selecting **Try information on this word** is effectively the same as selecting the word and selecting **Information** from the **Goodies** menu. **Information** gives a brief review of the term or statistic selected. It is discussed further in the last section of the **Help** chapter.

© 1989, SYSTAT, Inc

© 1989, SYSTAT, Inc

Index

ABS function, 69
Absolute value function, 69
ACS function, 69
Active window, 14
Addition (+), 69
Aggregated data, 88–90
[And] (see **Select...**)
Analysis, 122–139
[Analysis results] (see **Formats...**)
Analyzing subgroups, 134–36, 185
[And], 70
Arc hyperbolic tangent function, 69
Arccosine function, 69
Arcsine function, 69
Arctangent function, 69
Array, 52
ASN function, 69
ATH function, 69
ATN function, 69
[Autosave dialogs] (see **Formats...**)
AVG function, 72

Batches, 197
Beta distribution, 73
BCF function, 73
Blocks of data
  Changing, 60
  Selecting, 60
[Bold] (see **Formats...**)
BIF function, 73
BRN function, 73
Brushing tools, 174, 182–93
Bulleted cases, 186
Bundles, 130–33, 196
  Defining, 131
  Saving, 133
  Switching, 132

© 1989, SYSTAT, Inc.

By groups..., 134
[By groups off], 134
[Exclude missing], 135
[New page], 134

[Cancel], 129
Caret (^)
Cases, 46
Case weighting, 137–39
Cases, maximum number of, 50
Categorical variables, 166
CDF (see ZCF)
Changing blocks of data, 60
Changing data entries, 59
Changing variable names, 59
Character values, 54
Character variables, 47
Chi-square distribution, 73
[Clear], 129
Clear, 62
Clipboard (also see **Cut**, **Copy**, **Paste**)
Copying or cutting data into, 104
Placing output in, 120
Close box, 14
Color..., 157
Coloring objects, 162
Command batches, 197
Command help, 38–40
Command key abbreviations, 15, 197
Command logs, 197
Command interface, 4, 197
Copy, 62, 120
Copy graph, 120
Comments on output, 115
[Complex find...] (see **Find case...**)
Continuous variable, 166
[Condense] (see **Formats...**)
[Correlation] (see **Formats...**)
COS function, 69
Cosine function, 69

© 1989, SYSTAT, Inc.

[Covariance] (see **Formats...**)
Cross validation, 138
Cumulative functions,, 73
Cut, 62

Data Editor vs. Spreadsheet, 50
Data Editor, 44–66
Data files, 92–109
Data set, 146
Decimal places
  Shown in output, 125
  Shown in Editor, 82
[Decimals to show] (see **Formats...**)
Default, 25
Define bundles..., 131
Deleting cases, 60
Deleting variables, 60
Descriptive statistics, 24
Dialog boxes, 128–33
  Keyboard shortcuts, 128
  Options, 129
  Saving settings, 130
  Selecting functions, 128
  Selecting variables, 128
Dimension, 52
[Dissimilarity] (see **Formats...**)
Distribution functions, 73
Division (/), 69
[DLOG] (see **Save as...**)
Documentation, overview of, 9
Double precision, 96
Dragging objects, 142-22, 148, 159-61
Drawing tools, 158–63
Double precision, 96

ECF function,73
Editing data, 59–64
EIF function, 73
[Enter down] (see **Formats...**)

© 1989, SYSTAT, Inc.

[Enter right] (see **Formats...**)
Entering data, 48
Entering variable names, 51
Erlang variate, 74
Equal to (=), 70
ERN function, 73
Evaluation of expressions, 76
[Exclude missing] (see **By groups...**)
Exclusive option, 129
EXP function, 69
Exploratory data analysis (EDA), 164–93
Exponential distribution functions, 73
Exponential function, 69
Exponentiation (^), 69
Exporting data, 104–05
  Copying into the clipboard, 104
  Saving text files, 104
[Extend] (see **Formats...**)
[Extended results], 125
Extracting subsets of data, 186
Extract selected... , 84
Extreme value variate, 74
Eye, 174

F distribution, 73
FCF function, 73
FIF function, 73
Filenames, 94
Fill worksheet..., 76–77
Filling variables with random data, 77
Find case..., 78–81
  [Complex find...], 78
  [Test variable], 78
  [Variable or expression], 78
Find next, 78–81
Floppy disk installation, 5
Formats... (Data), 124–126
  [Analysis results], 124
  [Bold], 124
  [Condense], 124

© 1989, SYSTAT, Inc.

[Decimals to show], 125
[Extend], 124
[Extended results], 125
[Graphics], 125
[Italic], 124
[Outline], 124
[Menu click], 125
[Plain], 124
[Scroll analyses], 126
[Size], 126
[Shadow], 124
[Standard] results, 125
[Text font], 126
[Underline], 124
Formats... (Editor), 81–82
[Correlation], 82
[Covariance], 82
[Dissimilarity], 82
[Enter down], 82
[Enter right], 82
[Rectangular], 82
[Similarity], 82
[SSCP], 82
Format of numbers (see **Formats...**)
Frequency distributions, 73
FRN function, 73
Functions, 68–76

Gamma distribution, 73
Gamma variate, 73
GCF function, 73
GIF function, 73
Graph menu, 19–20
Graph options, 157
Color..., 157
Formats..., 157
Global settings..., 157
Graphic output, 110–20
Graphing, 140–163
Graphs, printing, 116
Greater than (>), 70

© 1989, SYSTAT, Inc.

Greater than or equal to (>=), 70
GRN function, 73
Grouping variables, 86–88

[Height] (see Ruler tool)
Help, 5, 28–33, 82, 196
  Command, 38–40
  Dialog, 30, 33–34
  Editor help, 37
  General, 30–33
  Instant, 30, 34–37
Hyperbolic arctangent, 69

Import..., 100–102
Importing data
  From Excel files, 102
  From map files, 103
  From text files, 100
  Pasting from the clipboard, 99–100
Inequalities, 70
Information, 41–43
Installing SYSTAT and SYGRAPH, 5
Instant graphs and statistics, 174, 197–201
  Example, 198–201
  Matrix cells, 198
  Variable name, 198
INT function, 69
Inverse functions, 73
Integer truncation function, 69
Inverse distribution functions, 73
[Italic] (see **Formats...**)

Keyboard shortcuts, 128

[Large screen] output, 117
Lasso tool (see Scatterplot brushing)
Left arrow tool (see Scatterplot brushing)
Less than (<), 70

© 1989, SYSTAT, Inc.

Less than or equal to (<=), 70
LGM function, 69
Linear, 167
LOG function, 69
Log gamma function (see LGM function)
Logical expressions, 70–72
Logical operators, 70–72

Macintosh basics, 13–15
Manipulating data files, 106–07
Map files, reading, 103
Marquee tool (see Scatterplot brushing)
Math..., 65–66
Matrix input, 82
MAX function, 72
Maximum file size, 50
Menu, 13
Menu bar, 13
[Menu click]
MIN function, 72
MIS function, 72
Missing values, 54–56
Moving graphs, 148
Movie..., 151–56
  Resolution, 151
  Time series example, 152–56
Multi-variable functions, 72
Multiplication (*), 69

Naming data files, 94
Naming variables, 51–53
Natural logarithm function, 69
Nested sort, 107
New page (see **By groups...**)
[Next plot], 151
Nonexclusive options, 129
Normal distribution, 73
[Normal] output, 117
[Not], 71
Not equal to (<>), 70

© 1989, SYSTAT, Inc.

Notes..., 113
NRAN (see ZRN)
Numeric variables, 47
Numerical values, 53

Object oriented, 158
Odometer, 66
One integrated environment, 4
Opening a SYSTAT data file, 18–19
Opening the Data Editor, 48–50
Operating tips, 194–202
Operators, 68–76
Options (**Graph** menu), 157
  Color..., 157
  Formats..., 157
  Global settings..., 157
Options, 129
  Exclusive, 129
  Nonexclusive, 129
  That take arguments, 129
[Or], 70
Order of expression evaluation, 76
Organization of SYSTAT, 99
Outliers, 170
[Outline] (see **Formats...**)
Output, 110–21
Output device control, 110–21
[Overlay plot], 150

Page formats..., 117
Page preview..., 116
Page Setup..., 98
Page titles, 117
Paging, 14
Paste, 63, 99–100
Pasting data from the clipboard, 99–100
  Character values, 99
  Delimiters, 99
  Variable names, 99

© 1989, SYSTAT, Inc.

PICT files, 112
Viewing, 112
[Plain] (see **Formats...**)
Plain text formats, 100
Plot tools and colors..., 158–63
Circle, 160
Line, 160
Oval, 159
Pointer, 161
Moving objects, 161
Removing objects, 161
Resizing objects, 161
Rectangle, 159
Rounded-corner rectangle, 159
Text, 160
Positioning graphs, 142–49
By mouse, 143
By typing, 144
Predictor variables
Categorical, 166
Continuous, 166
Print data..., 98, 119
Print graph..., 116-17
Print selected text..., 118
Printing a SYSTAT data file, 98–99
Printing output, 116–19

Random deviates, 74
Random distribution functions, 73
Random number generator, 75
Random subsample, 94
Rank..., 108
Ranking variables, 108
Large files, 108
Random variate functions, 73
Reading a text file, 100–102
Character values, 102
Delimiters, 101
Missing values, 102
Variable names, 101
Reading Excel files, 102–03

© 1989, SYSTAT, Inc.

Reading map files, 103
Reading SYSTAT files, 99
Recode..., 66–68
Record preferences, 127
[Rectangular] (see **Formats...**)
Redirecting output, 110–21
  Saving in files, 112
  Placing in clipboard, 120
  Printing, 116
Redo last analysis, 195
Regression, 26–27, 166
Relational operators, 70
Remove select, 84
Resize box, 148
Resizing graphs, 144
  By mouse, 144
  By typing, 144
Results to file as..., 114
Results to printer, 117, 118
Results to window, 114
RSEED function, 75
Round-off error
Ruler tool, 142
  [Height], 142
  Marquee, 142
  Resize box, 148
  [Show graph], 143
  [Width], 142
  [X origin], 144
  [Y origin], 144
  [Z height], 144

Save..., 56, 94
Save as..., 56, 95
  [SYSTAT], 95
  [Tabs], 95
  [Text], 95
  [Dlog], 95
Save graph as..., 112
Save selected text as..., 114

© 1989, SYSTAT, Inc.

Saved data, using, 115
Saving data, 56–58, 94–97, 112–115
Saving data in an textfile, 95, 104
  Delimiters, 105
Saving data produced by analyses, 115
Saving dialog boxes, 196
Scatterplot, 22–23
Scatterplot brushing, 182–83, 196
  Clicking on points, 183
  Lasso tool, 184
  Left arrow tool, 183
  Marquee tool, 184
  Selecting points, 183
  Tutorial, 187–93
[Scroll analyses] (see **Formats...**)
Scrolling, 14
Select..., 82–84, 135
Selections, turning off, 84
Selecting cases
  Algebraically, 82–84, 186
  Scatterplot brushing, 182–86
Selecting data, 84–85, (also see Scatterplot brushing)
  Blocks of data, 60
  Columns, 61
  Rows, 61
  Selecting subgroups, 135–36
[Shadow] (see **Formats...**)
Shortcuts, 195–202
Show graph (see Ruler tool)
Show options in effect..., 126
[Similarity] (see **Formats...**)
Simple data transformations, 64–77
SIN function, 69
Single precision, 96
Size box, 14
Sizing graphics, 142–49
Sort..., 106
Sorting data, 106–07
  Character order, 107
  Nested, 107
  Numeric order, 107
Sorting variables, 106

© 1989, SYSTAT, Inc.

Spreadsheet, 50
Space limits of procedures, 7
Special data structures, 86
Spin, 171–81
  Background, 173
  Brushing tools, 174
  Degrees, 173
  Eye, 174
  Home, 172
  Instant graphs, 174
  Labels, 173
  X,Y,Z, 173
  Tutorial, 174
Spinning, 5, 171–81
SPLOMs, 166
[Stack plot], 150
Standardize..., 109
Standard results, 125
Statistics, 24–27
Straightening scatterplots, 166–70
Stats menu, 24
Submit clipboard, 113
SQR function, 69
Square root function, 69
[SSCP] (see **Formats...**)
STD function, 72
Straightening scatterplots, 166–70
Subgroups
  Analyzing, 134–36
  Selecting, 135–36
Subscripted variables
Subset analysis, 185–86
Subtraction (–), 69
SUM function, 72
Supplements (see SYSTAT supplements)
[SYSTAT] (see **Save as...**)
SYSTAT data files, 7, 91–109
SYSTAT supplements, 8
  2SLS, 8
  DESIGN, 8
  EzPATH, 8
  LOGIT, 8

© 1989, SYSTAT, Inc.

PROBIT, 8
SETCOR, 8
SIGNAL, 8
SURVIVAL, 9
TESTAT, 8
TOBIT, 8

t distribution, 73
Tab-delimited, 95, 99, 101, 104–05
[Tabs] (see **Save as...**)
[Text] (see **Save as...**)
TAN function, 69
Tangent function, 69
TCF function, 73
Tear-off menu, 19–21
[Test variable] (see **Find case...**)
Text editor, 195
TIF function, 93
Text file
  Reading, 100–02
  Saving, 104
[Text font] (see **Formats...**)
Text list, 31
Text output, 111
Text window, 33
Thumb, 14
TIF function, 73
Time limits of procedures, 7
Title bar, 14
Title lines, 118
Transforming data, 65–77
Transpose..., 109
Transposing a file, 109
Triangular matrices, 90–91
Trigonometric functions, 69
TRN function, 73
Turning off selections, 84
Tutorial, 16–27

UCF function, 73

© 1989, SYSTAT, Inc.

UIF function, 73
Unary minus function, 69
[Underline] (see **Formats...**)
Uniform deviates, 73
Uniform distribution functions, 73
Uniform random numbers, 73
URN function , 73
[Use all variables] (see **Define bundles...**)
Using saved data, 115

Values, 53–56
  Character, 54
  Missing, 54
  Numeric, 53
Variable bundles, 130–33, 196
[Variable or expression] (see **Find case...**)
Variable names, 51–53
  Changing, 59
  Character, 51
  Entering, 51
  Subscripts, 51
Variables, 46
  Character, 47
  Numeric, 47
View window, 111
Viruses, 5

Weight..., 137
  [Weights off], 138
Weighting cases, 137–39
  Non-integer, 138
  Saving analyses after weighting, 138
Windows, 13
  As text editors, 113, 195
Worksheet, 46

[X origin] (see Ruler tool)
XCF function, 73
XIF function, 73

© 1989, SYSTAT, Inc.

XRN function, 73

[Y origin] (see Ruler tool)

[Z height] (see Ruler tool)
ZCF function, 73
ZIF function, 73
Zoom box, 14
ZRN function, 73

© 1989, SYSTAT, Inc.